Analisi Armonica e Timbro: La Trasformata di Fourier nella Percezione Sonora

Come la decomposizione in serie di Fourier spiega perché strumenti diversi suonano distintamente anche alla stessa frequenza fondamentale

Introduzione: Il problema del timbro nella fisica acustica

La questione fondamentale che questo saggio si propone di esplorare può essere formulata con estrema chiarezza: perché un La centrale (440 Hz) suonato da un violino risulta inconfondibilmente diverso dallo stesso La suonato da un clarinetto, da una tromba, o da un pianoforte, nonostante tutti questi strumenti producano vibrazioni con la stessa frequenza fondamentale? Questa domanda, apparentemente semplice, ha occupato fisici, matematici, fisiologi e psicologi per secoli, e la sua risposta richiede una comprensione profonda della natura matematica delle onde sonore e dei meccanismi di percezione uditiva.

La soluzione a questo enigma risiede in uno dei risultati più eleganti e potenti dell'analisi matematica: il teorema di Fourier, che afferma che qualsiasi funzione periodica sufficientemente regolare può essere rappresentata come somma infinita di funzioni sinusoidali di frequenze multiple della frequenza fondamentale. Questa decomposizione spettrale rivela che ciò che percepiamo come un singolo suono è in realtà una struttura complessa composta da molteplici componenti armoniche, ciascuna con la propria ampiezza e fase, e che è precisamente la distribuzione relativa di queste componenti armoniche a determinare il carattere timbrico distintivo di ciascuno strumento musicale.

Il presente saggio svilupperà sistematicamente la teoria matematica necessaria per comprendere questa fenomenologia, partendo dalle equazioni fondamentali del moto ondulatorio, procedendo attraverso la teoria della serie e della trasformata di Fourier, ed esaminando infine come queste strutture matematiche si manifestano concretamente nella produzione e percezione del suono musicale. L'esposizione integrerà rigore matematico con esempi concreti tratti dall'acustica degli strumenti musicali, dimostrando come la matematica pura illumini fenomeni percettivi immediati e quotidiani.

L'equazione d'onda e le sue soluzioni fondamentali

Il punto di partenza per qualsiasi trattamento quantitativo del suono è l'equazione d'onda, che descrive la propagazione di perturbazioni elastiche attraverso un mezzo. Consideriamo una perturbazione che si propaga in una dimensione spaziale (per esempio, lungo una

corda vibrante). Se denotiamo con y(x,t) lo spostamento verticale del punto della corda in posizione x al tempo t, l'equazione d'onda è:

$\partial^2 y/\partial t^2 = c^2 \, \partial^2 y/\partial x^2$

dove c è la velocità di propagazione dell'onda nel mezzo, determinata dalle proprietà elastiche del materiale. Questa equazione differenziale alle derivate parziali del secondo ordine ammette soluzioni di forma generale che possono essere rappresentate come sovrapposizione di onde sinusoidali.

La soluzione generale dell'equazione d'onda può essere scritta come combinazione lineare di due termini: un'onda che viaggia verso destra e un'onda che viaggia verso sinistra. Nel caso di una corda vibrante di lunghezza L fissata agli estremi (x = 0 e x = L), le condizioni al contorno impongono che y(0,t) = 0 e y(L,t) = 0 per ogni tempo t. Queste condizioni al contorno quantizzano le frequenze permesse, dando luogo ai modi normali di vibrazione.

I modi normali hanno la forma:

$y_n(x,t) = A_n \sin(n\pi x/L) \cos(\omega_n t + \varphi_n)$

dove n è un intero positivo (n = 1, 2, 3, ...), A_n è l'ampiezza del modo n-esimo, ω_n è la frequenza angolare data da:

$\omega_n = n\pi c/L$

e φ_n è la fase iniziale. La frequenza in Hertz corrispondente è:

$f_n = \omega_n/(2\pi) = nc/(2L)$

Questa espressione rivela che le frequenze permesse sono multipli interi della frequenza fondamentale $f_1 = c/(2L)$. Il modo n = 1 è chiamato fondamentale o prima armonica, mentre i modi con n > 1 sono chiamati armoniche superiori o overtones.

La soluzione generale per una corda vibrante è ottenuta sommando tutti i modi normali:

$y(x,t) = \Sigma(n=1 \text{ a } \infty) \, A_n \sin(n\pi x/L) \cos(\omega_n t + \varphi_n)$

Questa espressione è fondamentale: mostra che qualsiasi vibrazione della corda, per quanto complessa, può essere decomposta in una somma di vibrazioni sinusoidali semplici, ciascuna con la propria frequenza, ampiezza e fase. Le ampiezze A_n e le fasi φ_n sono determinate dalle condizioni iniziali specifiche (come la corda viene pizzicata, percossa, o eccitata con l'archetto).

Il teorema di Fourier per funzioni periodiche

Il risultato matematico che generalizza le osservazioni precedenti è il teorema di Fourier per funzioni periodiche. Consideriamo una funzione f(t) che è periodica con periodo T, ovvero f(t + T) = f(t) per ogni t. Se f(t) soddisfa certe condizioni di regolarità (essenzialmente, deve avere un numero finito di discontinuità e un numero finito di massimi e minimi in ogni

periodo, condizioni quasi sempre soddisfatte nelle applicazioni fisiche), allora può essere rappresentata come serie di Fourier:

$f(t) = a_0/2 + \Sigma(n=1 \text{ a } \infty) [a_n \cos(n\omega_0 t) + b_n \sin(n\omega_0 t)]$

dove $\omega_0 = 2\pi/T$ è la frequenza angolare fondamentale, e i coefficienti di Fourier a_n e b_n sono dati da:

$a_n = (2/T) \int_0^T f(t) \cos(n\omega_0 t) dt$

$b_n = (2/T) \int_0^T f(t) \sin(n\omega_0 t) dt$

Il termine $a_0/2$ rappresenta il valore medio della funzione sul periodo. Per segnali acustici, che hanno valore medio nullo (le perturbazioni di pressione oscillano attorno alla pressione atmosferica media), questo termine è generalmente zero.

Utilizzando l'identità trigonometrica:

$A \cos(\theta + \varphi) = A \cos(\varphi) \cos(\theta) - A \sin(\varphi) \sin(\theta)$

possiamo riscrivere la serie di Fourier in forma ampiezza-fase:

$f(t) = \Sigma(n=1 \text{ a } \infty) A_n \cos(n\omega_0 t + \varphi_n)$

dove:

$A_n = \sqrt{(a_n^2 + b_n^2)}$

$\varphi_n = -\arctan(b_n/a_n)$

Le quantità A_n rappresentano le ampiezze delle componenti armoniche, mentre φ_n rappresentano le loro fasi. Lo spettro di ampiezza $\{A_n\}$ fornisce informazione sulla distribuzione energetica tra le diverse frequenze, mentre lo spettro di fase $\{\varphi_n\}$ codifica le relazioni temporali tra le componenti.

Un risultato fondamentale della teoria di Fourier è il teorema di Parseval, che stabilisce la conservazione dell'energia:

$(1/T) \int_0^T |f(t)|^2 dt = (a_0^2/4) + (1/2) \Sigma(n=1 \text{ a } \infty) (a_n^2 + b_n^2) = (a_0^2/4) + (1/2) \Sigma(n=1 \text{ a } \infty) A_n^2$

Questo teorema afferma che l'energia totale del segnale (l'integrale del quadrato dell'ampiezza) è uguale alla somma delle energie delle singole componenti armoniche. Nel contesto acustico, dove l'intensità sonora è proporzionale al quadrato dell'ampiezza della perturbazione di pressione, il teorema di Parseval garantisce che l'intensità totale è la somma delle intensità delle componenti armoniche.

Rappresentazione in forma complessa

La serie di Fourier può essere espressa in forma particolarmente compatta utilizzando la notazione esponenziale complessa. Ricordando la formula di Eulero:

$e^{i\theta} = \cos(\theta) + i \sin(\theta)$

dove i è l'unità immaginaria ($i^2 = -1$), possiamo scrivere:

$\cos(\theta) = (e^{i\theta} + e^{-i\theta})/2$

$\sin(\theta) = (e^{i\theta} - e^{-i\theta})/(2i)$

Sostituendo queste espressioni nella serie di Fourier standard, otteniamo la serie di Fourier complessa:

$f(t) = \Sigma_{n=-\infty}^{+\infty} c_n e^{in\omega_0 t}$

dove i coefficienti complessi c_n sono dati da:

$c_n = (1/T) \int_0^T f(t) e^{-in\omega_0 t} dt$

I coefficienti c_n contengono sia informazione sull'ampiezza che sulla fase della componente n-esima. Specificatamente, se scriviamo c_n in forma polare:

$c_n = |c_n| e^{i\varphi_n}$

allora $|c_n|$ è l'ampiezza e φ_n è la fase della componente n-esima. La relazione con i coefficienti reali è:

$c_0 = a_0/2$

$c_n = (a_n - ib_n)/2$ per $n > 0$

$c_{-n} = (a_n + ib_n)/2 = c_n^*$ per $n > 0$

dove l'asterisco denota il coniugato complesso. Questa relazione $c_{-n} = c_n^*$ riflette il fatto che la funzione f(t) è reale.

La forma complessa della serie di Fourier è matematicamente elegante e spesso semplifica calcoli analitici. Tuttavia, per l'interpretazione fisica in termini di componenti armoniche udibili, la forma reale ampiezza-fase è generalmente più intuitiva.

Convergenza e regolarità

Una questione matematica importante riguarda le condizioni sotto cui la serie di Fourier converge effettivamente alla funzione originale. Il teorema di Dirichlet fornisce condizioni sufficienti: se f(t) è periodica, limitata, e ha un numero finito di discontinuità e di estremanti in ogni periodo, allora la serie di Fourier converge a f(t) in ogni punto di continuità, e converge al valore medio dei limiti destro e sinistro in ogni punto di discontinuità.

Per funzioni più regolari, la convergenza è più rapida. Se f(t) è continuamente differenziabile, i coefficienti di Fourier decadono come $1/n$ per n grande. Se f(t) ha derivate continue fino all'ordine k, i coefficienti decadono come $1/n^k$. Questo implica che funzioni più lisce richiedono meno termini della serie per essere approssimate accuratamente.

Nel contesto acustico, questa proprietà ha implicazioni percettive importanti. Forme d'onda con transitori bruschi (discontinuità nelle derivate) richiedono molte armoniche superiori per essere rappresentate accuratamente, risultando in un timbro "brillante" o "aspro". Forme d'onda più lisce, con contenuto armonico concentrato nelle frequenze più basse, producono timbri più "morbidi" o "scuri".

La trasformata di Fourier per segnali non periodici

Molti segnali acustici reali non sono strettamente periodici. Un violinista che suona una nota lunga produce un segnale che è approssimativamente periodico durante la fase di tenuta, ma include transitori non periodici all'attacco e al rilascio. Per trattare segnali generali, inclusi quelli non periodici, è necessario generalizzare la serie di Fourier alla trasformata di Fourier.

La trasformata di Fourier di una funzione f(t) definita sull'intera retta reale è:

$F(\omega) = \int_{-\infty}^{+\infty} f(t)\, e^{-i\omega t}\, dt$

dove ω è la frequenza angolare (misurata in radianti per secondo). La funzione F(ω) è chiamata spettro di Fourier o rappresentazione in frequenza di f(t). Generalmente F(ω) è una funzione complessa, e può essere scritta in forma polare:

$F(\omega) = |F(\omega)|\, e^{i\varphi(\omega)}$

dove |F(ω)| è lo spettro di ampiezza e φ(ω) è lo spettro di fase.

La trasformata di Fourier inversa permette di ricostruire f(t) da F(ω):

$f(t) = (1/(2\pi)) \int_{-\infty}^{+\infty} F(\omega)\, e^{i\omega t}\, d\omega$

Queste due equazioni formano una coppia di trasformazioni che permettono di passare tra la rappresentazione temporale e la rappresentazione spettrale del segnale.

Per segnali reali, vale la relazione F(-ω) = F(ω)*, che è l'analogo continuo della relazione $c_{-n} = c_n$* per le serie di Fourier. Questo implica che lo spettro di ampiezza è simmetrico: |F(-ω)| = |F(ω)|, mentre lo spettro di fase è antisimmetrico: φ(-ω) = -φ(ω).

Il teorema di Parseval per la trasformata di Fourier afferma:

$\int_{-\infty}^{+\infty} |f(t)|^2\, dt = (1/(2\pi)) \int_{-\infty}^{+\infty} |F(\omega)|^2\, d\omega$

garantendo la conservazione dell'energia nel passaggio dal dominio temporale al dominio frequenziale.

La trasformata di Fourier a tempo discreto

Nei sistemi digitali moderni di elaborazione del suono, i segnali sono campionati a intervalli temporali discreti. Se campioniamo un segnale f(t) a istanti $t_n = n\Delta t$, dove Δt è l'intervallo di campionamento e n è un intero, otteniamo una sequenza discreta $f_n = f(n\Delta t)$. La trasformata di Fourier a tempo discreto (DTFT) è definita come:

$F(\omega) = \Sigma_{n=-\infty \text{ a } +\infty} f_n\, e^{-i\omega n\Delta t}$

Questa è una funzione periodica in ω con periodo 2π/Δt. La trasformata inversa è:

f_n = (Δt/(2π)) ∫-π/Δt^(π/Δt) F(ω) e^(iωnΔt) dω

Il teorema di campionamento di Nyquist-Shannon stabilisce che un segnale a banda limitata (contenente solo frequenze fino a una frequenza massima f_max) può essere perfettamente ricostruito dai suoi campioni se la frequenza di campionamento f_s = 1/Δt soddisfa:

f_s ≥ 2f_max

La frequenza 2f_max è chiamata frequenza di Nyquist. Per l'audio digitale standard, f_s = 44100 Hz, permettendo di rappresentare frequenze fino a 22050 Hz, che eccede ampiamente il limite superiore dell'udito umano (circa 20000 Hz).

Nella pratica, si utilizza spesso la trasformata di Fourier discreta (DFT), che opera su sequenze finite di N campioni:

F_k = Σ(n=0 a N-1) f_n e^(-i2πkn/N)

con trasformata inversa:

f_n = (1/N) Σ(k=0 a N-1) F_k e^(i2πkn/N)

La DFT può essere calcolata efficientemente usando l'algoritmo FFT (Fast Fourier Transform), che richiede O(N log N) operazioni invece delle O(N²) operazioni richieste dal calcolo diretto, rendendo l'analisi spettrale computazionalmente fattibile per segnali lunghi.

La relazione tempo-frequenza e il principio di indeterminazione

Un aspetto fondamentale dell'analisi di Fourier con profonde implicazioni per l'acustica musicale è la relazione di indeterminazione tempo-frequenza, analoga al principio di indeterminazione di Heisenberg nella meccanica quantistica. Questa relazione stabilisce che non è possibile localizzare simultaneamente un segnale con precisione arbitraria sia nel dominio temporale che nel dominio frequenziale.

Formalmente, se definiamo la durata temporale efficace Δt di un segnale e la sua larghezza di banda efficace Δf mediante le deviazioni standard:

$(\Delta t)^2 = \int_{-\infty}^{+\infty} (t - t_0)^2 |f(t)|^2 dt / \int_{-\infty}^{+\infty} |f(t)|^2 dt$

$(\Delta f)^2 = \int_{-\infty}^{+\infty} (f - f_0)^2 |F(f)|^2 df / \int_{-\infty}^{+\infty} |F(f)|^2 df$

dove t_0 e f_0 sono i "centri" temporale e frequenziale del segnale, allora vale la relazione:

$\Delta t \cdot \Delta f \geq 1/(4\pi)$

Questa disuguaglianza implica che un segnale molto breve nel tempo (piccolo Δt) deve necessariamente avere un ampio contenuto spettrale (grande Δf), mentre un segnale quasi monocromatico (piccolo Δf) deve estendersi significativamente nel tempo (grande Δt).

Nel contesto musicale, questa relazione ha conseguenze importanti. Un suono percussivo molto breve, come il colpo di un triangolo o di piatti, non può avere una frequenza ben definita: il suo spettro si estende necessariamente su un ampio intervallo di frequenze. Viceversa, un suono con altezza molto precisa, come quello prodotto da un diapason, deve avere una durata ragionevolmente lunga per permettere al sistema uditivo di determinare accuratamente la frequenza.

Questa limitazione fondamentale influenza profondamente la percezione musicale. L'orecchio umano opera un compromesso tempo-frequenza: può localizzare eventi sonori con risoluzione temporale dell'ordine di millisecondi, ma questa risoluzione temporale fine comporta una risoluzione frequenziale relativamente grossolana. Per discriminare accuratamente piccole differenze di frequenza (come distinguere tra note musicali vicine), l'orecchio deve integrare l'informazione su durate temporali più lunghe.

L'energia spettrale e la densità spettrale di potenza

Per caratterizzare quantitativamente la distribuzione dell'energia tra le diverse frequenze, è utile introdurre il concetto di densità spettrale di potenza (PSD, Power Spectral Density). Per un segnale f(t) con trasformata di Fourier F(ω), la PSD è definita come:

$S(\omega) = |F(\omega)|^2$

Per segnali periodici rappresentati in serie di Fourier, la PSD consiste in impulsi delta posizionati alle frequenze armoniche:

$S(\omega) = \Sigma(n=-\infty \text{ a } +\infty) |c_n|^2 \delta(\omega - n\omega_0)$

dove δ è la funzione delta di Dirac. L'integrale della PSD su un intervallo di frequenze fornisce l'energia contenuta in quell'intervallo:

$E(\omega_1, \omega_2) = (1/(2\pi)) \int(\omega_1 \text{ a } \omega_2) S(\omega) d\omega$

Nel contesto acustico, la PSD determina la distribuzione dell'intensità sonora tra le diverse frequenze, ed è quindi direttamente correlata alla percezione timbrica. Due strumenti che suonano la stessa nota fondamentale ma hanno diverse PSD sonano diversamente perché l'energia è distribuita diversamente tra le armoniche.

Un parametro utile derivato dalla PSD è il centroide spettrale, definito come:

$f_c = (\int_0\text{^}(\infty) f S(f) df) / (\int_0\text{^}(\infty) S(f) df)$

dove abbiamo usato la frequenza ordinaria f invece della frequenza angolare ω (f = ω/(2π)). Il centroide spettrale rappresenta il "centro di gravità" dello spettro e correla fortemente con la percezione di "brillantezza" del timbro: suoni con centroide spettrale alto sono percepiti come brillanti, mentre suoni con centroide basso sono percepiti come scuri o opachi.

Altri descrittori spettrali utili includono la dispersione spettrale (spread), che misura quanto lo spettro è distribuito attorno al centroide, la skewness spettrale, che misura l'asimmetria della distribuzione spettrale, e la curtosi spettrale, che misura quanto la distribuzione è "piccata" o "piatta".

Funzioni finestra e trasformata di Fourier a finestra mobile

Nell'analisi di segnali musicali reali, che sono tipicamente non stazionari (le loro proprietà spettrali cambiano nel tempo), è necessario utilizzare tecniche che permettano di osservare come lo spettro evolve temporalmente. Questo si ottiene mediante la trasformata di Fourier a finestra mobile (STFT, Short-Time Fourier Transform).

L'idea fondamentale è moltiplicare il segnale per una funzione finestra w(t) che è concentrata attorno a un tempo particolare t_0, calcolare la trasformata di Fourier del segnale windowed, e poi spostare la finestra a diversi tempi. Formalmente:

$STFT(t_0, \omega) = \int_{-\infty}^{+\infty} f(t)\, w(t - t_0)\, e^{-i\omega t}\, dt$

La STFT fornisce una rappresentazione tempo-frequenza del segnale, mostrando quali frequenze sono presenti in quali momenti temporali.

La scelta della funzione finestra w(t) è critica e comporta un compromesso tra risoluzione temporale e risoluzione frequenziale, governato dal principio di indeterminazione discusso precedentemente. Finestre temporalmente strette forniscono buona risoluzione temporale ma scarsa risoluzione frequenziale, mentre finestre temporalmente larghe forniscono buona risoluzione frequenziale ma scarsa risoluzione temporale.

Diverse funzioni finestra sono comunemente utilizzate, ciascuna con proprietà specifiche:

Finestra rettangolare: $w(t) = 1$ per $|t| \leq T/2$, $w(t) = 0$ altrimenti

La finestra rettangolare ha la migliore risoluzione frequenziale ma presenta significativi "lobi laterali" nel dominio frequenziale (fenomeno di Gibbs), causando leakage spettrale.

Finestra di Hann: $w(t) = 0{,}5\,[1 + \cos(2\pi t/T)]$ per $|t| \leq T/2$, $w(t) = 0$ altrimenti

La finestra di Hann riduce il leakage spettrale rispetto alla finestra rettangolare, al costo di una risoluzione frequenziale leggermente peggiore.

Finestra di Hamming: $w(t) = 0{,}54 + 0{,}46 \cos(2\pi t/T)$ per $|t| \leq T/2$, $w(t) = 0$ altrimenti

Simile alla finestra di Hann ma con coefficienti ottimizzati per minimizzare ulteriormente i lobi laterali.

Finestra Gaussiana: $w(t) = e^{-t^2/(2\sigma^2)}$

La finestra Gaussiana è ottimale nel senso che minimizza il prodotto $\Delta t \cdot \Delta f$, saturando la relazione di indeterminazione.

La STFT è alla base dello spettrogramma, una rappresentazione visiva comunemente usata in acustica musicale dove il tempo è sull'asse orizzontale, la frequenza sull'asse verticale, e l'intensità di ciascuna componente frequenziale è rappresentata mediante colore o intensità di grigio.

Fisica della produzione sonora negli strumenti musicali

Meccanismi di eccitazione e risonanza

La produzione del suono in uno strumento musicale coinvolge generalmente due elementi fondamentali: un meccanismo di eccitazione che fornisce energia al sistema e un risonatore che seleziona e amplifica particolari frequenze. La natura di questi due elementi determina in modo cruciale le caratteristiche spettrali del suono prodotto e quindi il timbro dello strumento.

I meccanismi di eccitazione possono essere classificati in diverse categorie. Negli strumenti a corda (violini, chitarre, pianoforti), l'eccitazione avviene mediante pizzicamento (pizzicato), percussione (martelletti del pianoforte), o sfregamento con archetto. Negli strumenti a fiato, l'eccitazione è fornita dal flusso d'aria che induce oscillazioni autoeccitate attraverso vari meccanismi: il bordo tagliente negli strumenti a imboccatura naturale (flauti), le ance semplici o doppie negli strumenti ad ancia (clarinetti, oboi, fagotti), la vibrazione delle labbra negli ottoni. Negli strumenti a percussione, l'eccitazione è un impulso meccanico diretto.

Il risonatore può essere una corda vibrante, una colonna d'aria, una membrana, una piastra, o strutture più complesse. Ciascun tipo di risonatore ha modi normali di vibrazione con frequenze caratteristiche determinate dalla geometria e dalle proprietà elastiche del materiale.

Analisi dettagliata delle corde vibranti

Consideriamo una corda di lunghezza L, densità lineare μ (massa per unità di lunghezza), e tensione T. L'equazione d'onda per piccole perturbazioni trasversali y(x,t) è:

$$\mu \, \partial^2 y/\partial t^2 = T \, \partial^2 y/\partial x^2$$

che può essere riscritta come:

$$\partial^2 y/\partial t^2 = c^2 \, \partial^2 y/\partial x^2$$

dove la velocità di propagazione è $c = \sqrt{(T/\mu)}$.

Le condizioni al contorno $y(0,t) = y(L,t) = 0$ (corda fissata agli estremi) quantizzano i modi normali. Il modo n-esimo ha frequenza:

$$f_n = (n/(2L)) \sqrt{(T/\mu)} = n f_1$$

dove $f_1 = (1/(2L)) \sqrt{(T/\mu)}$ è la frequenza fondamentale. Questa relazione mostra che le frequenze armoniche sono esattamente multipli interi della fondamentale: $f_n = n f_1$.

La forma spaziale del modo n-esimo è:

$$\varphi_n(x) = \sin(n\pi x/L)$$

e la vibrazione completa è:

$$y_n(x,t) = A_n \sin(n\pi x/L) \cos(2\pi f_n t + \varphi_n)$$

La vibrazione generale della corda è la sovrapposizione di tutti i modi:

$y(x,t) = \Sigma(n=1 \text{ a } \infty) A_n \sin(n\pi x/L) \cos(2\pi f_n t + \varphi_n)$

Le ampiezze A_n e le fasi φ_n sono determinate dalle condizioni iniziali, ovvero dalla forma e dalla velocità della corda all'istante iniziale $t = 0$. Per una corda pizzicata in un punto x_0, la distribuzione iniziale di spostamento è triangolare:

$y(x,0) = h\, x/x_0$ per $0 \le x \le x_0$ $y(x,0) = h\,(L-x)/(L-x_0)$ per $x_0 < x \le L$

dove h è l'altezza massima dello spostamento. La velocità iniziale è nulla: $\partial y/\partial t(x,0) = 0$.

Applicando la decomposizione in serie di Fourier a questa distribuzione iniziale, si ottiene:

$A_n = (2hL^2)/(\pi^2 n^2 x_0(L-x_0)) \sin(n\pi x_0/L)$

$\varphi_n = 0$ per ogni n

Questa formula rivela diversi aspetti importanti:

1. L'ampiezza decade come $1/n^2$ per n grande, indicando che le armoniche superiori hanno ampiezze progressivamente minori.

2. Il fattore $\sin(n\pi x_0/L)$ causa la soppressione di particolari armoniche quando x_0 coincide con nodi di quei modi. Per esempio, se la corda è pizzicata esattamente al centro ($x_0 = L/2$), allora $\sin(n\pi/2) = 0$ per n pari, e le armoniche pari sono assenti. Questo spiega perché il suono di una chitarra o di un'arpa cambia significativamente a seconda di dove la corda viene pizzicata.

3. Pizzicando vicino a un estremo (x_0 piccolo o x_0 vicino a L), si eccitano fortemente molte armoniche superiori, producendo un timbro brillante. Pizzicando al centro, il suono è più morbido poiché le armoniche pari mancano.

Per una corda percossa (come nel pianoforte), l'analisi è più complessa poiché la condizione iniziale coinvolge sia spostamento che velocità. Un modello semplificato assume che il martelletto impartisca un impulso di velocità concentrato in un punto x_0:

$\partial y/\partial t(x,0) = v_0\, \delta(x - x_0)$

dove δ è la funzione delta di Dirac e v_0 è l'intensità dell'impulso. Questo porta a:

$A_n = (2v_0L)/(\pi^2 n^2 c) \sin(n\pi x_0/L)$

Il decadimento è ancora $1/n^2$, ma le fasi sono $\varphi_n = -\pi/2$ (corrispondente a un seno invece che a un coseno), riflettendo il fatto che l'eccitazione è di velocità piuttosto che di spostamento.

Smorzamento e decadimento armonico

Le corde reali non vibrano indefinitamente ma perdono energia gradualmente attraverso vari meccanismi di smorzamento: resistenza interna del materiale (viscoelasticità), perdite al

ponte di supporto, e resistenza dell'aria. Lo smorzamento può essere modellato aggiungendo un termine proporzionale alla velocità nell'equazione d'onda:

$\partial^2 y/\partial t^2 + 2\gamma\, \partial y/\partial t = c^2\, \partial^2 y/\partial x^2$

dove γ è il coefficiente di smorzamento. Le soluzioni sono ora oscillazioni smorzate esponenzialmente:

$y_n(x,t) = A_n\, e^{(-\gamma t)} \sin(n\pi x/L) \cos(2\pi f_n t + \varphi_n)$

dove le frequenze sono leggermente modificate dallo smorzamento:

$f_n \approx (n/(2L))\, \sqrt{(T/\mu)}\, \sqrt{(1 - \gamma^2/(4\pi^2 f_n^2))}$

Per smorzamenti deboli ($\gamma << 2\pi f_n$), la correzione di frequenza è trascurabile.

Un aspetto cruciale è che il coefficiente di smorzamento γ è generalmente diverso per armoniche diverse. In particolare, per molte corde reali, γ_n aumenta approssimativamente come n o n^2, implicando che le armoniche superiori decadono più rapidamente delle armoniche inferiori. Questo fenomeno ha conseguenze percettive importanti: durante il decadimento di una nota di pianoforte, il timbro diventa progressivamente più scuro poiché le componenti ad alta frequenza spariscono più rapidamente.

Il fattore di qualità Q di un modo risonante è definito come:

$Q_n = \pi f_n/\gamma_n$

e rappresenta il numero di oscillazioni necessarie perché l'ampiezza si riduca del fattore e (circa 2,718). Modi con Q alto risuonano per molto tempo, mentre modi con Q basso decadono rapidamente. Per un buon violino, Q può essere dell'ordine di 100-1000 per le armoniche fondamentali, mentre per percussioni smorzate come i timpani può essere molto minore.

Inarmonicità nelle corde reali

L'analisi precedente assume corde ideali perfettamente flessibili. Nella realtà, le corde hanno rigidità finita, specialmente le corde spesse e corte usate nel registro grave del pianoforte. La rigidità introduce un termine aggiuntivo nell'equazione d'onda:

$\partial^2 y/\partial t^2 = c^2\, \partial^2 y/\partial x^2 - \kappa^2\, \partial^4 y/\partial x^4$

dove $\kappa^2 = EI/(\mu)$, con E modulo di Young del materiale e I momento di inerzia della sezione trasversale.

Le frequenze dei modi normali diventano:

$f_n = (n/(2L))\, \sqrt{(T/\mu)}\, \sqrt{(1 + (n^2\pi^2\kappa^2)/(c^2L^2))}$

Espandendo per piccola rigidità:

$f_n \approx n\, f_1\, (1 + (n^2\pi^2\kappa^2)/(2c^2L^2))$

Il termine correttivo $n^2\pi^2\kappa^2/(2c^2L^2)$ causa una deviazione dalla relazione armonica perfetta. Le armoniche superiori hanno frequenze leggermente più alte rispetto ai multipli esatti della fondamentale. Questo fenomeno è chiamato inarmonicità.

Il coefficiente di inarmonicità B è definito come:

$B = (\pi^3\kappa^2)/(c^2L^2) = (\pi^3 EI)/(TL^4)$

e la frequenza del modo n-esimo è approssimata da:

$f_n \approx n f_1 \sqrt{(1 + Bn^2)} \approx n f_1 (1 + Bn^2/2)$

Per le corde del pianoforte, B varia tipicamente da circa 0,0001 per le corde lunghe e sottili del registro acuto fino a 0,01 o più per le corde corte e spesse del registro grave. Nel registro grave estremo, la settima armonica può essere alzata del 5-10% rispetto alla frequenza armonica ideale $7f_1$.

L'inarmonicità ha conseguenze importanti per la percezione dell'altezza e per l'accordatura. Nel pianoforte, le ottave sono tipicamente "stirate" leggermente (le ottave superiori sono accordate leggermente più alte e le ottave inferiori leggermente più basse rispetto all'accordatura teorica) per compensare parzialmente gli effetti dell'inarmonicità sulla percezione dell'altezza.

Colonne d'aria e strumenti a fiato

Gli strumenti a fiato utilizzano colonne d'aria come risonatori. L'equazione d'onda per piccole perturbazioni di pressione p(x,t) in una colonna d'aria è:

$\partial^2 p/\partial t^2 = c^2 \partial^2 p/\partial x^2$

dove $c \approx 343$ m/s è la velocità del suono in aria a temperatura ambiente.

Le condizioni al contorno dipendono dalla configurazione degli estremi. Un estremo aperto (in comunicazione con l'atmosfera) è approssimativamente un nodo di pressione ($p \approx 0$), mentre un estremo chiuso è un antinodo di pressione ($\partial p/\partial x = 0$).

Per un tubo cilindrico aperto a entrambi gli estremi di lunghezza L, le frequenze dei modi normali sono:

$f_n = (nc)/(2L) = n f_1$

identiche a quelle di una corda vibrante. Il tubo aperto produce una serie armonica completa.

Per un tubo cilindrico chiuso a un estremo e aperto all'altro (come in alcuni flauti o negli organi a canne chiuse), le condizioni al contorno richiedono un antinodo all'estremo chiuso e un nodo all'estremo aperto. Le frequenze sono:

$f_n = ((2n-1)c)/(4L) = (2n-1) f_1$

dove $n = 1, 2, 3, \ldots$ Questo tubo produce solo le armoniche dispari: f_1, $3f_1$, $5f_1$, $7f_1$, ...

Questa differenza fondamentale spiega perché un clarinetto (che si comporta approssimativamente come un tubo cilindrico chiuso, poiché l'ancia funge da estremo chiuso) ha un timbro marcatamente diverso da un flauto o da un oboe (che si comportano come tubi aperti): il clarinetto manca delle armoniche pari, conferendogli un carattere sonoro distintivo, specialmente nel registro basso chiamato "registro chalumeau".

I tubi conici (come oboe, fagotto, sassofono) hanno comportamenti più complessi. Un cono tronco aperto a entrambi gli estremi produce una serie armonica completa come un cilindro aperto, ma con distribuzioni spaziali di pressione diverse. Questo influenza l'efficienza con cui diverse armoniche sono eccitate e radiate.

Correzione di fine e lunghezza efficace

Nella realtà, le condizioni al contorno agli estremi aperti di un tubo non sono perfettamente semplici. L'aria all'estremità di un tubo aperto non si arresta esattamente al bordo geometrico, ma "continua" leggermente oltre, creando un effetto di "correzione di fine" (end correction). La lunghezza acustica efficace è:

$$L_{eff} = L + \Delta l$$

dove Δl è la correzione di fine, tipicamente dell'ordine di 0,6 volte il raggio del tubo per un'estremità flangiata, e circa 0,3 volte il raggio per un'estremità non flangiata.

Questa correzione diventa significativa per strumenti con tubi larghi e corti, come certi registri dell'organo. La frequenza fondamentale effettiva è:

$$f_1 = c/(2L_{eff}) = c/(2(L + \Delta l))$$

leggermente inferiore a quanto previsto dalla lunghezza geometrica sola.

Inoltre, la correzione di fine dipende debolmente dalla frequenza, introducendo piccole inarmonicità anche nelle colonne d'aria. Tuttavia, queste inarmonicità sono generalmente molto minori rispetto a quelle causate dalla rigidità delle corde.

Meccanismi di eccitazione non lineari

Molti strumenti a fiato coinvolgono meccanismi di eccitazione intrinsecamente non lineari. Per esempio, in un clarinetto, l'ancia si comporta come una valvola che controlla il flusso d'aria in modo non lineare. Quando la pressione nella bocca del musicista supera la pressione nella colonna d'aria sottostante, l'ancia si apre permettendo il passaggio dell'aria. Questo flusso riduce la pressione (effetto Bernoulli), causando la chiusura dell'ancia. Il ciclo si ripete a una frequenza determinata dalla risonanza della colonna d'aria.

Questo meccanismo può essere modellato mediante un'equazione non lineare che accoppia la dinamica dell'ancia con le oscillazioni di pressione nella colonna. Una versione semplificata è:

$$dp/dt + (c^2/V)\, Q = (c^2/V)\, S\, u(\Delta p)$$

dove p è la pressione nella colonna, Q è il flusso volumetrico, V è il volume, S è l'area della sezione trasversale, u(Δp) è la velocità del flusso attraverso l'apertura dell'ancia (funzione non lineare della differenza di pressione Δp tra bocca e colonna), e c è la velocità del suono.

La non linearità in u(Δp) causa la generazione di armoniche superiori anche quando la colonna d'aria ha una singola frequenza risonante dominante. Questo è un meccanismo cruciale per la produzione del timbro caratteristico degli strumenti ad ancia.

Negli ottoni, la non linearità proviene dalla vibrazione delle labbra del musicista, che funzionano come ance doppie. Le labbra hanno una frequenza propria di vibrazione che può essere variata dal musicista mediante tensione muscolare. L'interazione non lineare tra la vibrazione delle labbra e le risonanze della colonna d'aria determina la frequenza di suono risultante e il contenuto armonico.

L'archetto del violino: eccitazione stick-slip

Il meccanismo mediante cui l'archetto eccita le corde del violino (e altri strumenti ad arco) è particolarmente complesso e affascinante. Quando l'archetto si muove sulla corda, quest'ultima è trascinata dall'attrito ("stick") fino a che la forza elastica della corda eccede la forza di attrito, causando uno scorrimento rapido ("slip"). Questo ciclo stick-slip si ripete periodicamente alla frequenza fondamentale della corda.

Durante la fase stick, la corda si muove con l'archetto e si deforma elasticamente. Durante la fase slip, la corda scorre velocemente rispetto all'archetto, rilasciando l'energia elastica accumulata. Questo genera un impulso di velocità che si propaga lungo la corda come un'onda.

Un modello semplificato è il modello di Raman, che considera il movimento in un punto della corda dove agisce l'archetto. La forza di attrito ha una caratteristica non lineare: l'attrito statico è maggiore dell'attrito dinamico. Se denotiamo con μ_s il coefficiente di attrito statico e con μ_d il coefficiente di attrito dinamico, la forza di attrito è approssimativamente:

F = μ_s F_N quando la corda e l'archetto sono in fase stick (v_rel = 0) F = μ_d F_N sign(v_rel) quando sono in fase slip (v_rel ≠ 0)

dove F_N è la forza normale dell'archetto sulla corda e v_rel è la velocità relativa.

Questa non linearità nell'attrito è essenziale per il funzionamento dell'archetto. Helmholtz dimostrò che il movimento risultante della corda sotto l'azione dell'archetto è caratterizzato da una forma d'onda specifica chiamata "movimento di Helmholtz": la corda ha una forma a V che ruota attorno al suo perimetro. In un punto lontano dall'archetto, la velocità della corda è una funzione approssimativamente periodica composta da segmenti costanti collegati da transizioni brusche.

La trasformata di Fourier di questa forma d'onda "a dente di sega" ha ampiezze armoniche che decadono come 1/n:

A_n ∝ 1/n

Questo decadimento 1/n (più lento del decadimento $1/n^2$ per le corde pizzicate) genera uno spettro ricco di armoniche superiori, contribuendo al timbro caratteristico degli strumenti ad arco.

La posizione dell'archetto sulla corda influenza significativamente il timbro. Suonando vicino al ponte (sul ponticello), si eccitano fortemente le armoniche superiori, producendo un suono brillante e metallico. Suonando sulla tastiera (sul tasto), le armoniche superiori sono attenuate, producendo un suono più dolce e flautato.

Risonanze del corpo e formanti

La cassa di risonanza degli strumenti a corda (violini, chitarre, pianoforti) e il corpo degli strumenti a fiato non sono risonatori passivi ma hanno proprie frequenze di risonanza chiamate formanti che modificano significativamente lo spettro del suono prodotto.

Nel violino, la tavola armonica (top plate) e il fondo (back plate) hanno modi di vibrazione complessi con frequenze risonanti tipicamente nell'intervallo 200-3000 Hz. Quando una corda vibra, trasmette energia al ponte, che a sua volta eccita le vibrazioni del corpo. Le componenti spettrali vicine alle risonanze del corpo sono amplificate, mentre altre sono attenuate.

Le risonanze più importanti nel violino includono:

- Il modo A0 (circa 270 Hz): modo di respirazione dove l'aria nel corpo vibra in fase attraverso le f-holes.
- Il modo B1- (circa 470 Hz): primo modo della tavola armonica, con antinodo al centro.
- Il modo B1+ (circa 540 Hz): primo modo del fondo.

Queste risonanze creano picchi nella curva di risposta in frequenza dello strumento. La presenza e la posizione di questi picchi contribuiscono significativamente alla qualità timbrica distintiva di ciascun violino.

Un buon violino ha risonanze ben distribuite in frequenza e con fattori Q appropriati (né troppo alti, che produrrebbero colorazione eccessiva, né troppo bassi, che non fornirebbero amplificazione sufficiente). La costruzione del violino, inclusi lo spessore delle tavole, la forma e posizione delle f-holes, la posizione della barra armonica interna (bass bar) e dell'anima (sound post), è ottimizzata empiricamente attraverso secoli di sviluppo per ottenere distribuzioni di risonanze desiderabili.

Negli strumenti a fiato, le risonanze del corpo (campana degli ottoni, corpo dei legni) influenzano similmente il timbro. La campana di una tromba, per esempio, ha la funzione cruciale di accoppiare efficacemente l'onda acustica nella colonna d'aria con l'aria esterna, facilitando la radiazione delle frequenze alte. La forma della campana influenza quali armoniche sono radiate efficacemente e quindi il bilanciamento spettrale percepito.

Transitori di attacco e decadimento

Un aspetto cruciale del timbro che non è catturato dall'analisi dello stato stazionario è la struttura temporale del suono, particolarmente i transitori all'attacco (come il suono inizia) e al decadimento (come il suono finisce).

Il transitorio di attacco contiene informazione percettivamente molto saliente che permette di identificare strumenti anche quando la porzione stazionaria del suono è mascherata. Esperimenti psicoacustici hanno dimostrato che se si rimuove il primo breve segmento (primi 50-100 ms) di una registrazione di uno strumento musicale, l'identificazione dello strumento diventa significativamente più difficile, anche se la porzione stazionaria è preservata.

I transitori hanno generalmente contenuto spettrale molto ampio e non stazionario. Nell'attacco di una nota di pianoforte, per esempio, il martelletto colpisce inizialmente la corda producendo un rumore impulsivo ad ampio spettro, seguito dalla costruzione graduale delle risonanze della corda e del corpo. Le diverse armoniche raggiungono le loro ampiezze stazionarie con diverse costanti di tempo, creando una struttura temporale complessa.

Un modello semplice per l'inviluppo temporale dell'ampiezza è il modello ADSR (Attack, Decay, Sustain, Release):

- Attack: fase iniziale dove l'ampiezza cresce da zero al massimo
- Decay: fase dove l'ampiezza decresce dal massimo al livello di sostegno
- Sustain: fase di ampiezza approssimativamente costante durante cui la nota è mantenuta
- Release: fase finale dove l'ampiezza decade a zero quando la nota è rilasciata

Matematicamente, questo può essere rappresentato come:

$A(t) = A_{attack}(t)$ per $0 \leq t < t_a$ $A(t) = A_{decay}(t)$ per $t_a \leq t < t_d$ $A(t) = A_{sustain}$ per $t_d \leq t < t_s$ $A(t) = A_{release}(t)$ per $t \geq t_s$

dove A_{attack}, A_{decay}, e $A_{release}$ sono tipicamente funzioni esponenziali o polinomiali che interpolano tra i livelli appropriati.

In realtà, diverse armoniche hanno inviluppi ADSR diversi. Nel pianoforte, le armoniche superiori decadono più rapidamente durante la fase di release, causando un progressivo scurimento del timbro durante il decadimento della nota.

Fisiologia e psicoacustica della percezione del timbro

Anatomia e funzionamento dell'orecchio umano

Per comprendere come la decomposizione spettrale analizzata matematicamente si manifesta nella percezione effettiva, è necessario esaminare la fisiologia dell'orecchio umano, che opera essenzialmente come un analizzatore spettrale biologico.

L'orecchio è tradizionalmente diviso in tre sezioni: orecchio esterno, medio e interno. L'orecchio esterno comprende il padiglione auricolare e il canale uditivo esterno, che termina alla membrana timpanica (timpano). Il padiglione auricolare ha una geometria complessa che introduce filtraggi dipendenti dalla direzione, contribuendo alla localizzazione spaziale

dei suoni. Il canale uditivo, con lunghezza tipica di circa 2,5 cm, agisce come un risonatore tubolare con frequenza di risonanza attorno a 3000 Hz, amplificando le frequenze in questa regione.

L'orecchio medio contiene tre ossicini (martello, incudine, staffa) che formano una catena di leve che trasmette le vibrazioni dal timpano alla finestra ovale della coclea. Questo sistema di leve fornisce adattamento di impedenza tra l'aria e i fluidi della coclea, massimizzando il trasferimento di energia acustica.

L'orecchio interno contiene la coclea, una struttura a spirale riempita di fluido dove avviene l'analisi spettrale vera e propria. La coclea umana compie circa 2,75 giri con lunghezza totale sviluppata di circa 35 mm. La membrana basilare, una struttura elastica che corre lungo tutta la lunghezza della coclea, è fondamentale per l'analisi frequenziale.

La coclea come analizzatore spettrale meccanico

La membrana basilare ha proprietà meccaniche che variano gradualmente lungo la sua lunghezza. Vicino alla finestra ovale (base), è stretta (circa 0,1 mm) e rigida. All'apice, è più larga (circa 0,5 mm) e più flessibile. Questa gradazione delle proprietà meccaniche crea una mappatura spaziale delle frequenze: suoni ad alta frequenza producono massima vibrazione vicino alla base, mentre suoni a bassa frequenza producono massima vibrazione verso l'apice.

Il fisiologo Georg von Békésy, premio Nobel per questa ricerca, dimostrò sperimentalmente che quando una frequenza pura eccita la coclea, si forma un'onda viaggiante sulla membrana basilare. Questa onda cresce in ampiezza mentre viaggia verso l'apice, raggiunge un massimo a una posizione caratteristica dipendente dalla frequenza, e poi decade rapidamente.

La posizione x del massimo di vibrazione sulla membrana basilare per una frequenza f può essere approssimata da una relazione logaritmica:

$$x(f) = L - a \log_{10}(f/f_0)$$

dove $L \approx 35$ mm è la lunghezza totale della coclea, $a \approx 8$ mm, e $f_0 \approx 20000$ Hz. Questa relazione implica che intervalli musicali costanti (come ottave) corrispondono a distanze approssimativamente costanti sulla membrana basilare, una proprietà chiamata organizzazione tonotopica.

L'inviluppo dell'onda viaggiante per una frequenza f ha approssimativamente forma Gaussiana attorno alla posizione caratteristica:

$$A(x, f) = A_0 \exp[-(x - x(f))^2/(2\sigma^2)]$$

dove σ determina la larghezza della risposta. Questa forma Gaussiana implica che ogni punto della membrana basilare risponde non a una singola frequenza ma a un intervallo di frequenze, con massima sensibilità alla frequenza caratteristica di quel punto.

Le cellule ciliate e la trasduzione meccanoelettrica

Sulla superficie della membrana basilare si trova l'organo del Corti, che contiene le cellule sensoriali: circa 3500 cellule ciliate interne (IHC) e 12000 cellule ciliate esterne (OHC), organizzate in file parallele lungo la coclea.

Le cellule ciliate interne sono i principali trasduttori sensoriali. Quando la membrana basilare vibra, le stereociglia sulla superficie delle IHC sono deflesse dalla membrana tettoria sovrastante. Questa deflessione meccanica apre canali ionici meccano-sensibili nelle stereociglia, permettendo l'ingresso di ioni potassio che depolarizzano la cellula. La depolarizzazione causa il rilascio di neurotrasmettitori che eccitano le fibre del nervo uditivo.

Le cellule ciliate esterne hanno una funzione diversa e straordinaria: sono elettromotili, ovvero cambiano lunghezza in risposta alla depolarizzazione. Questo meccanismo di feedback positivo amplifica meccanicamente le vibrazioni della membrana basilare, aumentando la sensibilità e la selettività frequenziale dell'orecchio. La disfunzione delle OHC è una causa comune di perdita uditiva.

L'amplificazione cocleare fornita dalle OHC è maggiore per suoni deboli e diminuisce per suoni forti, un comportamento non lineare chiamato compressione cocleare. Questo comportamento può essere modellato matematicamente con una funzione di input-output compressiva:

Output = Input^α

dove α ≈ 0,3 per suoni di intensità moderata. Questa compressione contribuisce alla vasta gamma dinamica dell'udito umano (circa 120 dB, corrispondente a un rapporto di intensità di un milione di milioni).

I filtri uditivi e le bande critiche

La risposta in frequenza di ciascuna posizione sulla membrana basilare può essere modellata come un filtro passa-banda. Questi "filtri uditivi" hanno larghezza di banda che aumenta con la frequenza centrale.

La larghezza di banda del filtro uditivo centrato alla frequenza f è tradizionalmente misurata come banda critica (critical bandwidth), definita come la larghezza di banda entro cui l'integrazione energetica è approssimativamente lineare. La banda critica CBW(f) può essere approssimata da:

CBW(f) = 24,7 (4,37 f + 1) Hz

per f in kHz, oppure da formule più semplici come:

CBW(f) ≈ 0,69 f^(0,69) Hz

Per frequenze basse (f < 500 Hz), la banda critica è approssimativamente costante a circa 100 Hz. Per frequenze alte, è approssimativamente proporzionale alla frequenza: CBW(f) ≈ 0,2 f.

Una scala percettivamente più rilevante è la scala Bark, dove l'intera gamma uditiva (da circa 20 Hz a 20 kHz) è divisa in 24 bande critiche numerate da 0 a 24 Bark. La relazione approssimata tra frequenza f (in Hz) e posizione z (in Bark) è:

$$z = 13 \arctan(0{,}00076\, f) + 3{,}5 \arctan((f/7500)^2)$$

oppure, inversamente:

$$f \approx 600 \sinh(z/6)$$

La scala Bark riflette la risoluzione frequenziale effettiva dell'orecchio: due suoni puri separati da meno di una banda critica interagiscono percettivamente (mascheramento, rugosità), mentre suoni separati da più di una banda critica sono percepiti come relativamente indipendenti.

Una scala correlata è la scala ERB (Equivalent Rectangular Bandwidth), definita come la larghezza del filtro rettangolare che ha la stessa area del filtro uditivo reale con la stessa altezza di picco. La scala ERB-rate è:

$$\text{ERB-rate} = 21{,}4 \log_{10}(1 + 0{,}00437\, f)$$

dove f è in Hz. Questa scala fornisce una mappatura leggermente diversa ma concettualmente simile alla scala Bark.

La codifica neurale dell'informazione spettrale

L'informazione spettrale estratta dalla coclea è trasmessa al cervello attraverso circa 30000 fibre del nervo uditivo. Esistono due meccanismi principali mediante cui l'informazione frequenziale è codificata:

Codifica di luogo (place coding): Fibre diverse innervano diverse posizioni sulla membrana basilare e quindi rispondono preferenzialmente a diverse frequenze. La frequenza di un suono è codificata dalla popolazione di fibre che risponde più vigorosamente. Questo meccanismo è efficace per tutte le frequenze udibili.

Codifica temporale (temporal coding): Per frequenze basse (approssimativamente sotto 4-5 kHz), le fibre del nervo uditivo possono "phase-lock" al segnale acustico, ovvero sparare potenziali d'azione in fase con la forma d'onda. Sebbene una singola fibra non possa sparare a ogni ciclo per frequenze sopra circa 1 kHz (a causa del periodo refrattario), la popolazione collettiva di fibre può rappresentare la struttura temporale fine del segnale.

Per suoni complessi contenenti multiple componenti armoniche, entrambi i meccanismi contribuiscono alla rappresentazione. Il pattern spaziale di attivazione sulla membrana basilare (e quindi sulla popolazione di fibre neurali) riflette lo spettro del suono. Simultaneamente, per le componenti a bassa frequenza, il pattern temporale di scarica neurale conserva informazione sulla struttura temporale fine.

Percezione dell'altezza e teoria del pattern temporale

La percezione dell'altezza (pitch) di suoni complessi è un fenomeno psicoacustico affascinante che illustra come il sistema uditivo integra l'informazione spettrale e temporale.

Quando un suono contiene armoniche multiple della frequenza fondamentale f_0 (come nelle note musicali tipiche), percepiamo un'altezza corrispondente a f_0 anche se la fondamentale fisica è assente o molto debole. Questo fenomeno è chiamato "missing fundamental" o "virtual pitch".

Esistono due teorie principali per spiegare la percezione della virtual pitch:

Teoria del pattern spettrale (place theory): Il cervello identifica armoniche individuali attraverso i picchi di attivazione sulla membrana basilare, riconosce che queste frequenze sono multipli interi di una frequenza comune, e inferisce questa frequenza fondamentale. Matematicamente, questo corrisponde a trovare il massimo comun divisore delle frequenze armoniche.

Se le frequenze misurate sono $f_1 = n_1 f_0$, $f_2 = n_2 f_0$, ..., f_k = n_kf_0 dove n_1, n_2, ..., n_k sono interi, allora:

f_0 = GCD(f_1, f_2, ..., f_k)

In pratica, le frequenze sono misurate con errore finito, quindi un algoritmo robusto cerca la frequenza f_0 che minimizza gli errori residui:

f_0 = argmin_f Σ_i min_n [f_i - nf]²

Teoria del pattern temporale (temporal theory): La forma d'onda di un suono con armoniche multiple ha una struttura temporale periodica con periodo $T_0 = 1/f_0$. Il sistema uditivo può estrarre questa periodicità dal pattern temporale di scarica neurale, anche quando le singole armoniche non sono risolvibili.

Matematicamente, questo corrisponde a calcolare l'autocorrelazione del segnale neurale:

$R(\tau) = \int s(t)\, s(t + \tau)\, dt$

I picchi dell'autocorrelazione si verificano a multipli del periodo T_0, permettendo l'estrazione della fondamentale.

Esperimenti indicano che entrambi i meccanismi contribuiscono, con peso relativo dipendente dalle caratteristiche del suono. Per suoni con molte armoniche risolte a bassa frequenza, il pattern temporale domina. Per suoni con armoniche non risolte ad alta frequenza, il pattern spettrale è più importante.

Mascheramento spettrale e temporale

Il mascheramento è il fenomeno per cui la presenza di un suono (masker) riduce la percezione di un altro suono (target). Il mascheramento è cruciale per comprendere come i diversi componenti spettrali di suoni complessi interagiscono percettivamente.

Mascheramento spettrale: Un suono puro di frequenza f_m e intensità I_m maschera altri suoni puri in un intervallo di frequenze attorno a f_m. La soglia di mascheramento (la minima intensità necessaria per il target per essere percepibile in presenza del masker) dipende dalla separazione frequenziale tra masker e target.

Vicino alla frequenza del masker (entro la banda critica), la soglia di mascheramento è alta. A frequenze lontane, il mascheramento diminuisce rapidamente. Asimmetricamente, il mascheramento verso frequenze superiori è maggiore del mascheramento verso frequenze inferiori. Un masker a 1000 Hz maschera più efficacemente target a 1100 Hz che target a 900 Hz.

Il pattern di mascheramento può essere modellato matematicamente. Se il masker ha frequenza f_m e livello di pressione sonora L_m (in dB), la soglia di mascheramento per un target a frequenza f_t è approssimativamente:

L_t(f_t) = L_m + S(f_t/f_m)

dove S(x) è una funzione che descrive la forma della curva di mascheramento. Una forma approssimata è:

S(x) = -24 - 230/x + 0,2 x dB per x < 1 S(x) = -24 + 0,2 x dB per x ≥ 1

dove x = f_t/f_m è il rapporto di frequenze.

Mascheramento temporale: Il mascheramento si verifica anche nel dominio temporale. Un masker intenso può mascherare un target che lo precede (pre-mascheramento) o che lo segue (post-mascheramento). Il post-mascheramento può persistere per 50-200 ms dopo la cessazione del masker, mentre il pre-mascheramento è più breve (circa 20 ms).

Il mascheramento temporale è modellato con funzioni che decadono esponenzialmente o come potenze del tempo:

$M(t) = M_0 \exp(-t/\tau)$ per post-mascheramento

dove M_0 è l'intensità del mascheramento all'istante finale del masker e τ è la costante di tempo (tipicamente 50-100 ms).

Rugosità sensoriale e battimenti

Quando due suoni puri di frequenze vicine f_1 e f_2 sono presentati simultaneamente, l'interferenza tra di loro produce fenomeni percettivi dipendenti dalla separazione frequenziale.

Per separazioni molto piccole ($\Delta f = |f_2 - f_1| < 15$-$20$ Hz), percepiamo battimenti (beats): modulazioni periodiche dell'ampiezza con frequenza pari a Δf. La forma d'onda risultante è:

$s(t) = A_1 \cos(2\pi f_1 t) + A_2 \cos(2\pi f_2 t) = 2A \cos(2\pi((f_1+f_2)/2)t) \cos(2\pi((f_1-f_2)/2)t)$

per $A_1 = A_2 = A$. L'inviluppo modulato $\cos(2\pi \Delta f\, t/2)$ produce la sensazione di battimenti.

Per separazioni intermedie (approssimativamente 15-75 Hz), la modulazione è troppo rapida per essere percepita come battimenti distinti, ma produce una sensazione di rugosità (roughness) o dissonanza sensoriale. La rugosità è massima per $\Delta f \approx 30$-40 Hz (circa un quarto di banda critica) e diminuisce per separazioni maggiori.

La rugosità può essere quantificata mediante modelli psicoacustici. Un modello semplice propone che la rugosità R tra due componenti con frequenze f_1, f_2 e ampiezze A_1, A_2 è:

$$R(f_1, f_2, A_1, A_2) \propto A_1 A_2 (\Delta f) \exp(-\kappa \Delta f)$$

dove κ è una costante che determina il decadimento con la separazione frequenziale. Il prodotto $A_1 A_2$ riflette il fatto che la rugosità richiede entrambe le componenti siano udibili; il fattore Δf riflette l'aumento iniziale con la separazione; il fattore esponenziale riflette la diminuzione per grandi separazioni.

Per suoni complessi con molte componenti armoniche, la rugosità totale è approssimativamente la somma delle rugosità pairwise tra tutte le coppie di componenti che cadono entro la stessa banda critica.

Consonanza e dissonanza: basi psicoacustiche

La sensazione di consonanza o dissonanza tra suoni è parzialmente spiegata dalla rugosità sensoriale. Intervalli musicali consonanti (ottave, quinte, terze maggiori) hanno armoniche che o coincidono esattamente o sono sufficientemente separate da evitare rugosità. Intervalli dissonanti (seconde minori, tritoni) hanno molte armoniche vicine che producono rugosità.

Consideriamo due note con frequenze fondamentali f_1 e f_2 e armoniche $n_1 f_1$ e $n_2 f_2$. Le armoniche coincidono quando:

$$n_1 f_1 = n_2 f_2$$

ovvero quando:

$$f_2/f_1 = n_1/n_2$$

Rapporti di frequenze esprimibili con piccoli interi (come 2/1 per l'ottava, 3/2 per la quinta, 4/3 per la quarta, 5/4 per la terza maggiore) producono molte coincidenze armoniche e quindi consonanza.

Herman Helmholtz nel diciannovesimo secolo propose che la consonanza deriva dall'assenza di rugosità tra armoniche. Calcolò modelli di consonanza basati sulla rugosità totale tra tutte le coppie di armoniche. Questi modelli predicono correttamente che intervalli con rapporti di piccoli interi sono più consonanti.

Tuttavia, la consonanza musicale è anche influenzata da fattori culturali e contestuali oltre alla psicoacustica pura. L'addestramento musicale modifica la percezione di consonanza/dissonanza, e l'evoluzione storica della musica occidentale ha progressivamente accettato come consonanti intervalli precedentemente considerati dissonanti.

Colorazione timbrica e formalismo HRTF

Il timbro percepito di un suono dipende non solo dalle sue proprietà spettrali intrinseche ma anche dal filtraggio imposto dal canale di trasmissione, incluso l'ambiente acustico e l'anatomia dell'ascoltatore.

La Head-Related Transfer Function (HRTF) descrive il filtraggio imposto dalla testa, dal torso e dai padiglioni auricolari. L'HRTF è diversa per suoni provenienti da direzioni diverse e fornisce le cues principali per la localizzazione spaziale.

Matematicamente, se s(t) è il segnale alla sorgente e s_ear(t) è il segnale al timpano, la relazione nel dominio frequenziale è:

$S_{ear}(\omega, \theta, \varphi) = H(\omega, \theta, \varphi)\, S(\omega)$

dove $H(\omega, \theta, \varphi)$ è l'HRTF dipendente dalla frequenza ω e dagli angoli azimutali ed elevazionali (θ, φ) della sorgente.

L'HRTF introduce notches (tagli) e picchi nello spettro che dipendono dalla direzione. Per esempio, suoni provenienti dall'alto producono notches caratteristici attorno a 8-10 kHz a causa della riflessione dal bordo del padiglione auricolare. Questi notches aiutano a disambiguare l'elevazione.

Nel contesto musicale, l'HRTF è importante per la spazializzazione audio e per la sintesi binaurale, dove si cerca di creare l'impressione di sorgenti sonore posizionate in direzioni specifiche mediante opportuno filtraggio del segnale.

Integrazione temporale e risoluzione spettrale

L'orecchio non opera su istanti infinitesimali ma integra informazione su finestre temporali finite. La durata della finestra di integrazione temporale influenza la risoluzione spettrale secondo la relazione di indeterminazione tempo-frequenza discussa precedentemente.

Per stimare accuratamente la frequenza di un suono puro, l'orecchio deve osservare il suono per una durata sufficiente. La risoluzione frequenziale Δf migliora con l'aumentare della durata di osservazione T secondo:

$\Delta f \propto 1/T$

Per durate molto brevi (< 10 ms), la discriminazione di frequenza è scarsa. Per durate di 100-200 ms, si raggiunge la risoluzione massima, che è circa 0,2-0,3% della frequenza centrale per ascoltatori addestrati (corrispondente a circa 1/30 di semitono).

Questa integrazione temporale implica che transitori molto brevi non possono avere altezza ben definita. Un click impulsivo è percepito come un rumore ad ampio spettro piuttosto che come una nota con altezza definita, consistentemente con il fatto che la trasformata di Fourier di un impulso $\delta(t)$ è una costante (spettro uniforme a tutte le frequenze).

Per suoni musicali tipici, l'altezza emerge dopo circa 30-50 ms di osservazione, mentre la piena caratterizzazione timbrica richiede spesso l'intero transitorio di attacco (100-200 ms).

Questo spiega perché la rimozione dell'attacco rende difficile l'identificazione dello strumento anche quando la porzione stazionaria è preservata.

Analisi spettrale comparativa di strumenti musicali

Metodologia dell'analisi spettrale strumentale

Per comprendere concretamente come la teoria di Fourier spiega le differenze timbriche tra strumenti, esaminiamo analisi spettrali dettagliate di vari strumenti che suonano la stessa nota fondamentale. Utilizzeremo il La4 (A4) a 440 Hz come riferimento standard, sebbene i principi si applichino a qualsiasi frequenza.

L'analisi tipica procede mediante registrazione digitale del suono a frequenza di campionamento elevata (tipicamente 44,1 kHz o superiore), seguita da applicazione della FFT a segmenti successivi del segnale per ottenere l'evoluzione temporale dello spettro. La lunghezza della finestra FFT rappresenta un compromesso tra risoluzione temporale e frequenziale: finestre di 2048-4096 campioni (circa 46-93 ms a 44,1 kHz) forniscono risoluzione frequenziale di circa 10-20 Hz, adeguata per risolvere armoniche individuali nelle frequenze musicali tipiche.

Spettro del diapason: il caso quasi ideale

Il diapason rappresenta il caso più semplice, producendo un suono quasi puro dominato dalla frequenza fondamentale con armoniche superiori molto deboli. Per un diapason che suona La4 (440 Hz), l'analisi spettrale rivela:

- Fondamentale (440 Hz): ampiezza relativa 1,00 (riferimento)
- Seconda armonica (880 Hz): ampiezza relativa 0,01-0,03
- Terza armonica (1320 Hz): ampiezza relativa < 0,01
- Armoniche superiori: essenzialmente assenti

Il rapporto segnale-rumore spettrale (SSNR), definito come:

$SSNR = 10 \log_{10}(P_{fondamentale} / \Sigma P_{armoniche})$

dove P denota la potenza (proporzionale al quadrato dell'ampiezza), è tipicamente superiore a 20 dB per un diapason, indicando predominanza quasi totale della fondamentale.

Il decadimento temporale del diapason è approssimativamente esponenziale:

$A(t) = A_0 e^{(-t/\tau)}$

con costante di tempo τ tipicamente nell'intervallo 5-10 secondi per diapason di buona qualità. La purezza spettrale del diapason deriva dalla simmetria geometrica e dalla modalità di eccitazione (percussione sui rebbi paralleli che vibrano in opposizione di fase).

Spettro del flauto traverso: predominanza della fondamentale con armoniche controllate

Il flauto traverso, strumento ad imboccatura naturale, produce uno spettro relativamente semplice dominato dalle armoniche inferiori. Per un La4 suonato con dinamica mezzoforte, l'analisi tipica rivela:

- Fondamentale (440 Hz): 1,00
- Seconda armonica (880 Hz): 0,15-0,30
- Terza armonica (1320 Hz): 0,10-0,20
- Quarta armonica (1760 Hz): 0,05-0,15
- Quinta armonica (2200 Hz): 0,03-0,10
- Armoniche superiori (n ≥ 6): < 0,05

Lo spettro mostra decadimento monotono delle ampiezze armoniche approssimativamente esponenziale:

$A_n \approx A_1 e^{(-\alpha n)}$

con α ≈ 0,3-0,5 dipendente dall'intensità di esecuzione e dal registro.

Un aspetto distintivo del flauto è la forte dipendenza dello spettro dalla dinamica. Suonato pianissimo, il suono è quasi puro come un diapason. Suonato fortissimo, le armoniche superiori diventano significativamente più prominenti, con la seconda armonica che può raggiungere 40-50% dell'ampiezza della fondamentale.

La distribuzione spettrale può essere caratterizzata dal centroide spettrale:

$f_c = (\Sigma\, n f_1 A_n^2) / (\Sigma\, A_n^2)$

Per il flauto in mezzoforte, $f_c \approx 600$-800 Hz, relativamente vicino alla fondamentale, coerentemente con il carattere timbrico percepito come "puro" o "dolce".

La struttura temporale del suono di flauto è caratterizzata da un attacco relativamente lento (50-150 ms) dove le armoniche si costruiscono gradualmente, seguito da una fase stazionaria dove lo spettro è approssimativamente stabile, e un decadimento rapido (< 50 ms) quando il musicista interrompe il flusso d'aria.

Spettro del clarinetto: dominanza delle armoniche dispari

Il clarinetto, con il suo tubo cilindrico chiuso a un'estremità, produce teoricamente solo armoniche dispari, sebbene in pratica armoniche pari deboli siano presenti a causa di non idealità. Per un La4 in registro chalumeau (registro grave), l'analisi rivela:

- Fondamentale (440 Hz): 1,00
- Seconda armonica (880 Hz): 0,05-0,10
- Terza armonica (1320 Hz): 0,60-0,90
- Quarta armonica (1760 Hz): 0,02-0,05
- Quinta armonica (2200 Hz): 0,30-0,50
- Sesta armonica (2640 Hz): < 0,02
- Settima armonica (3080 Hz): 0,15-0,30

Il rapporto A_3/A_2 (ampiezza terza armonica / ampiezza seconda armonica) è tipicamente 10-20 per il clarinetto, mentre per la maggior parte degli altri strumenti a fiato è < 2. Questa predominanza delle armoniche dispari conferisce al clarinetto il suo carattere timbrico distintivo, particolarmente evidente nel registro grave.

La soppressione delle armoniche pari può essere quantificata mediante l'indice di disparità:

D = (Σ_dispari A_n²) / (Σ_pari A_n²)

Per il clarinetto in registro grave, D ≈ 10-50, mentre per la maggior parte degli altri strumenti D ≈ 1-2.

Nel registro superiore (clarinetto), l'effetto è meno pronunciato poiché la nota fondamentale percepita corrisponde alla terza armonica del tubo (il clarinetto "overblow" alla dodicesima piuttosto che all'ottava), e le armoniche pari della fondamentale percepita corrispondono ad armoniche che sono multipli di sei della fondamentale acustica del tubo.

Il transitorio di attacco del clarinetto è caratterizzato da costruzione progressiva delle armoniche con diverse costanti di tempo. La fondamentale raggiunge tipicamente il massimo prima della terza armonica, creando una variazione temporale del colore timbrico durante l'attacco.

Spettro dell'oboe: ricchezza armonica e formanti prominenti

L'oboe, strumento ad ancia doppia con tubo conico, produce uno spettro particolarmente ricco con molte armoniche prominenti. Per un La4:

- Fondamentale (440 Hz): 1,00
- Seconda armonica (880 Hz): 0,80-1,20
- Terza armonica (1320 Hz): 0,60-1,00
- Quarta armonica (1760 Hz): 0,40-0,70
- Quinta armonica (2200 Hz): 0,50-0,90
- Sesta armonica (2640 Hz): 0,30-0,60
- Settima armonica (3080 Hz): 0,40-0,80
- Ottava armonica (3520 Hz): 0,20-0,50
- Armoniche 9-15: 0,10-0,40 ciascuna

Notevole è la presenza di armoniche che possono superare in ampiezza la fondamentale. Particolarmente la seconda e la terza armonica sono spesso estremamente forti. Questo crea uno spettro molto denso con centroide spettrale alto:

f_c ≈ 1200-1800 Hz

molto superiore alla fondamentale.

L'oboe presenta formanti prominenti (risonanze del corpo dello strumento) attorno a 1100-1300 Hz e 2500-3000 Hz. Queste formanti amplificano selettivamente armoniche che cadono in queste regioni frequenziali, indipendentemente dalla nota suonata. Per il La4, la terza armonica (1320 Hz) cade vicino alla prima formante e viene significativamente amplificata.

La presenza di formanti crea una caratteristica distintiva dello spettro dell'oboe: quando la nota suonata cambia, armoniche diverse vengono amplificate dalle formanti fisse, ma le posizioni frequenziali delle formanti rimangono approssimativamente costanti. Questo può essere visualizzato in uno spettrogramma dove le formanti appaiono come bande orizzontali di maggiore intensità che attraversano diverse note.

Il transitorio di attacco dell'oboe è complesso, con un "chirp" iniziale (rapida discesa frequenziale) seguito dalla stabilizzazione sulla frequenza target. Questo transitorio, che dura tipicamente 30-80 ms, contiene energia significativa sopra 5000 Hz ed è cruciale per l'identificazione dello strumento.

Spettro della tromba: brillantezza e armoniche acute

La tromba, rappresentante degli ottoni, produce uno spettro ricco esteso fino a frequenze molto alte. Per un La4 suonato con dinamica forte:

- Fondamentale (440 Hz): 1,00
- Seconda armonica (880 Hz): 0,60-0,90
- Terza armonica (1320 Hz): 0,70-1,10
- Quarta armonica (1760 Hz): 0,50-0,80
- Quinta armonica (2200 Hz): 0,60-1,00
- Sesta armonica (2640 Hz): 0,40-0,70
- Settima armonica (3080 Hz): 0,50-0,90
- Ottava armonica (3520 Hz): 0,30-0,60
- Armoniche 9-20: significative (0,10-0,50)

Lo spettro della tromba si estende tipicamente fino alla 20-30esima armonica con ampiezze ancora percettivamente rilevanti. Per La4, questo significa componenti significative fino a 8800-13200 Hz.

Il decadimento delle ampiezze armoniche è più lento che per gli strumenti a fiato di legno, approssimativamente:

$A_n \approx A_1 \, n^{(-\beta)}$

con $\beta \approx 0{,}5\text{-}0{,}8$ (decadimento power-law piuttosto che esponenziale).

La campana della tromba introduce formanti attorno a 1200, 1800 e 2500 Hz (valori approssimativi variabili con lo strumento specifico). Queste formanti contribuiscono alla brillantezza caratteristica del timbro della tromba.

Un fenomeno particolare della tromba (e altri ottoni) è la forte dipendenza dello spettro dall'intensità dinamica e dalla tecnica di esecuzione. Suonata pianissimo con embouchure rilassata, la tromba può produrre uno spettro relativamente semplice simile a quello di un corno. Suonata fortissimo con embouchure tesa, lo spettro diventa estremamente ricco con armoniche alte prominenti.

La non linearità dell'eccitazione (vibrazione delle labbra) introduce anche armoniche secondarie deboli (non multiple intere della fondamentale) e sidebands di modulazione, contribuendo alla complessità timbrica.

Spettro del violino: complessità armonica e variabilità temporale

Il violino produce forse lo spettro più complesso e variabile temporalmente tra gli strumenti comuni. Per un La4 suonato con arco sul ponticello (brillante):

- Fondamentale (440 Hz): 1,00
- Seconda armonica (880 Hz): 0,50-0,80
- Terza armonica (1320 Hz): 0,40-0,70
- Quarta armonica (1760 Hz): 0,60-1,00
- Quinta armonica (2200 Hz): 0,50-0,90
- Sesta armonica (2640 Hz): 0,40-0,70
- Settima-decima armonica: 0,30-0,60 ciascuna
- Armoniche superiori (11-30): 0,05-0,30

Suonato sul tasto (dolce), le ampiezze delle armoniche alte sono significativamente ridotte:

- Armoniche 1-3: simili al caso precedente
- Armoniche 4-6: ridotte del 30-50%
- Armoniche > 6: ridotte del 50-80%

Il corpo del violino introduce formanti complesse, le più importanti delle quali sono:

- A0 ("air resonance"): 270-290 Hz
- B1- ("corpus mode"): 450-480 Hz
- B1+ ("corpus mode"): 520-550 Hz

Queste risonanze creano picchi e valli nello spettro che dipendono dalla sovrapposizione tra armoniche della nota suonata e frequenze di risonanza del corpo.

Una caratteristica distintiva del violino è la variabilità temporale fine dello spettro anche durante la fase nominalmente stazionaria. L'azione dell'arco produce oscillazioni periodiche nell'ampiezza e nella frequenza (vibrato se intenzionale, ma anche fluttuazioni involontarie), creando sidebands attorno a ciascuna armonica:

$f_{n,\text{sidebands}} = n f_0 \pm k f_v$

dove $f_v \approx 5\text{-}8$ Hz è la frequenza del vibrato e $k = 1, 2, 3, \ldots$

Inoltre, il meccanismo stick-slip dell'arco produce jitter di frequenza (variazioni irregolari della frequenza istantanea) quantificabili mediante:

$$\text{Jitter} = \sqrt{\langle (f(t) - \langle f \rangle)^2 \rangle} / \langle f \rangle$$

dove $\langle\rangle$ denota media temporale. Per un violino ben suonato, Jitter ≈ 0,001-0,003 (0,1-0,3%).

Il transitorio di attacco del violino dipende criticamente dalla velocità e pressione dell'arco. Un attacco rapido e forte produce un transitorio ricco con energia significativa sopra 10 kHz nelle prime decine di millisecondi. Un attacco lento e morbido produce costruzione graduale delle armoniche con meno contenuto ad alta frequenza.

Spettro del pianoforte: inarmonicità e decadimento differenziale

Il pianoforte presenta caratteristiche spettrali uniche derivanti dalla rigidità delle corde e dall'accoppiamento con la tavola armonica. Per un La4 (corda centrale del pianoforte):

Immediatamente dopo l'attacco:

- Fondamentale (440 Hz): 1,00
- Seconda armonica (882 Hz): 0,40-0,60
- Terza armonica (1326 Hz): 0,30-0,50
- Quarta armonica (1772 Hz): 0,25-0,45
- Quinta armonica (2220 Hz): 0,20-0,40

Notare le frequenze leggermente inarmoniche (882 Hz invece di 880 Hz, ecc.) dovute alla rigidità della corda discussa precedentemente.

Un secondo dopo l'attacco (durante il decadimento):

- Fondamentale: 1,00 (normalizzata)
- Seconda armonica: 0,20-0,35
- Terza armonica: 0,10-0,20
- Armoniche superiori: significativamente ridotte

Il decadimento differenziale può essere modellato con:

$A_n(t) = A_n(0)\, e^{(-\gamma_n t)}$

dove $\gamma_n \approx \gamma_1 (1 + \beta(n-1))$ con $\beta \approx 0{,}1\text{-}0{,}3$, implicando che la seconda armonica decade circa 10-30% più rapidamente della fondamentale, la terza circa 20-60% più rapidamente, ecc.

Questo decadimento differenziale causa una deriva timbrica durante la nota: il suono diventa progressivamente più scuro (centroide spettrale diminuisce) durante il decadimento. Questa evoluzione timbrica è parte integrante del carattere espressivo del pianoforte.

L'inarmonicità influenza anche la percezione dell'accordatura. Per compensare, i pianoforti sono tipicamente accordati con "ottave stirate":

$f_n^{(accordato)} \approx 2^{(n)} f_1 (1 + \varepsilon n)$

dove $\varepsilon \approx 0{,}001\text{-}0{,}003$ dipende dallo strumento specifico.

La tavola armonica del pianoforte introduce una curva di radiazione complessa, con efficienza che varia con la frequenza. Tipicamente, l'efficienza è bassa sotto 100 Hz, cresce verso un massimo attorno a 1000-2000 Hz, e poi diminuisce gradualmente verso frequenze alte. Questo filtraggio modifica lo spettro effettivamente radiato rispetto alle vibrazioni delle corde.

Spettro della voce umana: il caso più complesso

La voce umana rappresenta forse il caso più complesso di produzione sonora, combinando un'eccitazione periodica (vibrazione delle corde vocali) con un complesso sistema di risonanze (tratto vocale).

Le corde vocali vibrano periodicamente producendo un'onda di flusso d'aria pulsante (glottal flow). La forma d'onda del flusso glottico approssima una forma triangolare asimmetrica, con chiusura rapida e apertura più lenta. La trasformata di Fourier di questa forma d'onda produce uno spettro ricco:

$A_n \approx A_1 / n^{\alpha}$

con $\alpha \approx 1\text{-}1{,}5$ ("rolloff spettrale" di circa -12 a -18 dB per ottava).

Il tratto vocale (faringe, cavità orale, cavità nasale) agisce come un sistema di filtri risonanti. Le risonanze principali sono chiamate formanti e sono cruciali per distinguere diverse vocali. Per un maschio adulto che canta una vocale "a":

- F1 (prima formante): 650-750 Hz
- F2 (seconda formante): 1100-1300 Hz
- F3 (terza formante): 2400-2800 Hz
- F4 (quarta formante): 3200-3600 Hz

Queste formanti amplificano selettivamente armoniche della fondamentale che cadono vicino alle loro frequenze. Per un La4 (440 Hz) cantato con vocale "a":

- L'armonica 2 (880 Hz) cade vicino a F1 ed è amplificata
- L'armonica 3 (1320 Hz) cade vicino a F2 ed è fortemente amplificata
- Armoniche 5-6 (2200-2640 Hz) cadono vicino a F3

Lo spettro risultante mostra picchi pronunciati alle armoniche vicine alle formanti, creando la caratteristica struttura spettrale distintiva della vocale.

Cambiando la vocale (modificando la forma del tratto vocale), le frequenze formantiche cambiano mentre la frequenza fondamentale può rimanere costante. Per esempio, per la vocale "i":

- F1: 280-350 Hz
- F2: 2200-2400 Hz
- F3: 2900-3200 Hz

Questo sposta drasticamente quali armoniche sono amplificate, cambiando radicalmente il timbro percepito pur mantenendo la stessa altezza.

La voce presenta anche modulazioni fini:

Vibrato: oscillazione periodica della frequenza fondamentale, tipicamente 5-7 Hz con escursione di ±0,5-1,5 semitoni.

Shimmer: variazioni irregolari dell'ampiezza ciclo-per-ciclo, quantificato come:

Shimmer = (<|A_i - A_(i+1)|>) / <A_i>

dove A_i è l'ampiezza del ciclo i-esimo. Per voci normali, Shimmer < 0,05 (5%).

Jitter: variazioni irregolari del periodo ciclo-per-ciclo:

Jitter = (<|T_i - T_(i+1)|>) / <T_i>

dove T_i è il periodo del ciclo i-esimo. Per voci normali, Jitter < 0,01 (1%).

Voci patologiche possono avere shimmer e jitter molto più alti, contribuendo alla percezione di raucedine.

Sintesi del suono, elaborazione digitale e applicazioni moderne

Sintesi additiva: ricostruzione del suono mediante componenti armoniche

La comprensione della decomposizione spettrale mediante trasformata di Fourier conduce naturalmente alla sintesi additiva, dove suoni complessi sono generati sommando componenti sinusoidali individuali. Questo metodo costituisce l'implementazione diretta del teorema di Fourier: se un suono può essere decomposto in componenti armoniche, può essere ricostruito sommando quelle componenti.

Il segnale sintetizzato è espresso come:

$s(t) = \Sigma(n=1 \text{ a } N)\ A_n(t) \sin(2\pi n f_0 t + \varphi_n(t))$

dove $A_n(t)$ e $\varphi_n(t)$ sono l'ampiezza e la fase temporalmente variabili della n-esima armonica. La dipendenza temporale delle ampiezze permette di modellare i transitori di attacco e decadimento, mentre la dipendenza temporale delle fasi permette di modellare fenomeni come il vibrato.

Per esempio, per sintetizzare un suono di clarinetto, potremmo utilizzare:

- Numero di armoniche: N = 15
- Ampiezze iniziali: $A_n(0) = 0$ per n pari, $A_n(0) = A_1/n$ per n dispari
- Inviluppo di ampiezza: $A_n(t) = A_n(0)\ E(t)$ dove E(t) è una funzione ADSR
- Modulazione di frequenza (vibrato): $f_0(t) = 440\ [1 + 0{,}01 \sin(2\pi\ 6\ t)]$

Implementando questo algoritmo con parametri accuratamente scelti basati su analisi di registrazioni reali, è possibile generare suoni sintetici che sono percettivamente molto simili allo strumento target.

Tuttavia, la sintesi additiva pura presenta limitazioni. Catturare accuratamente tutti i dettagli del timbro di uno strumento reale richiede modellare non solo le ampiezze medie delle armoniche ma anche le loro fluttuazioni temporali fini, le correlazioni di fase tra armoniche, e i transitori complessi. Per un violino, per esempio, potrebbero essere necessarie 30-50

armoniche con inviluppi temporali e modulazioni individualmente specificati, risultando in un grande numero di parametri.

Nonostante queste complessità, la sintesi additiva è stata utilizzata estesamente, particolarmente negli organi elettronici Hammond (che utilizzano ruote tonali rotanti per generare armoniche individuali che sono poi sommate mediante drawbars controllabili dal musicista) e in sintetizzatori digitali moderni.

Sintesi sottrattiva: filtraggio di forme d'onda ricche

Un approccio complementare è la sintesi sottrattiva, dove si parte da una forma d'onda ricca di contenuto armonico (tipicamente un'onda a dente di sega o un'onda quadra) e si applicano filtri per rimuovere o attenuare selettivamente componenti frequenziali.

Un'onda a dente di sega ha armoniche con ampiezze:

$A_n = A_1/n$

fornendo uno spettro ricco con tutte le armoniche. Un'onda quadra ha solo armoniche dispari:

$A_n = A_1/n$ per n dispari, $A_n = 0$ per n pari

Applicando un filtro passa-basso con frequenza di taglio f_c, le armoniche sopra f_c sono attenuate. La funzione di trasferimento di un filtro passa-basso del primo ordine è:

$H(f) = 1 / \sqrt{(1 + (f/f_c)^2)}$

con rolloff di -6 dB per ottava sopra f_c. Filtri di ordine superiore hanno rolloff più ripidi: un filtro del secondo ordine ha -12 dB per ottava, un filtro del quarto ordine ha -24 dB per ottava.

Filtri risonanti, che hanno un picco a una frequenza specifica prima di attenuare, sono particolarmente utili per emulare le formanti degli strumenti. Un filtro risonante del secondo ordine ha funzione di trasferimento:

$H(f) = 1 / \sqrt{[(1 - (f/f_r)^2)^2 + (f/(f_r\ Q))^2]}$

dove f_r è la frequenza risonante e Q è il fattore di qualità. Per Q >> 1, il filtro ha un picco pronunciato a f_r con larghezza di banda $\Delta f \approx f_r/Q$.

Variando temporalmente i parametri del filtro (frequency sweeping), si possono creare effetti timbrici dinamici. Questo principio è alla base dei sintetizzatori sottrattivi classici come il Moog e l'ARP, dove oscillatori generano forme d'onda basilari che sono poi elaborate mediante filtri voltage-controlled.

La sintesi sottrattiva è computazionalmente più efficiente della sintesi additiva (richiede generare una singola forma d'onda e applicare alcuni filtri invece di generare e sommare decine di sinusoidi) ma offre meno controllo dettagliato dello spettro.

Sintesi mediante modulazione di frequenza (FM)

La sintesi FM, resa celebre dal sintetizzatore Yamaha DX7, sfrutta il fatto che la modulazione di frequenza di una sinusoide pura genera un ricco spettro di armoniche. Se modifichiamo istantaneamente la frequenza di un oscillatore (carrier) mediante un altro oscillatore (modulator), il segnale risultante è:

$s(t) = A \sin[2\pi f_c t + I \sin(2\pi f_m t)]$

dove f_c è la frequenza carrier, f_m è la frequenza modulator, e I è l'indice di modulazione.

Espandendo questa espressione mediante funzioni di Bessel, si ottiene:

$s(t) = A \Sigma(n=-\infty \text{ a } +\infty) J_n(I) \sin[2\pi(f_c + n f_m) t]$

dove $J_n(I)$ sono le funzioni di Bessel di prima specie di ordine n.

Questo mostra che la FM genera componenti spettrali (sidebands) alle frequenze $f_c + n f_m$ per tutti gli interi n. L'ampiezza della n-esima sideband è proporzionale a $J_n(I)$.

Quando il rapporto f_c/f_m è un numero razionale, le sidebands coincidono con armoniche di una frequenza fondamentale comune, producendo suoni armonici utilizzabili musicalmente. Per esempio, con $f_m = f_c$ (rapporto 1:1), le sidebands si verificano a frequenze 0, f_c, $2f_c$, $3f_c$, ..., producendo uno spettro armonico con fondamentale f_c.

L'indice di modulazione I controlla la ricchezza spettrale. Per I piccolo ($I << 1$):

$s(t) \approx A \sin(2\pi f_c t) + (AI/2)[\cos(2\pi(f_c - f_m)t) - \cos(2\pi(f_c + f_m)t)]$

contenente solo la carrier e le due sidebands del primo ordine. Per I grande, molte sidebands sono significative, creando uno spettro molto ricco.

Variando temporalmente I, si possono creare transitori di attacco brillanti seguiti da toni più morbidi, emulando l'evoluzione spettrale di molti strumenti acustici. La sintesi FM è particolarmente efficace per emulare suoni metallici (campane, vibrafono, pianoforte elettrico) e certi strumenti a fiato.

Configurazioni più complesse utilizzano multiple carriers e modulators in topologie elaborate (algoritmi FM), permettendo grande varietà timbrica con relativamente pochi parametri.

Sintesi granulare e sintesi basata su tabelle d'onda

La sintesi granulare decompone il suono in brevi segmenti (grani) di durata tipica 10-100 ms, ciascuno dei quali è una porzione windowed di una forma d'onda. Il suono completo è ricostruito sovrapponendo molti grani con diverse temporizzazioni, altezze, e ampiezze.

Matematicamente, un grano è:

$g_i(t) = w(t - t_i) s(\alpha_i(t - t_i))$

dove w(t) è una funzione finestra (tipicamente Gaussiana o Hann), t_i è il tempo di inizio del grano, s(·) è la forma d'onda sorgente, e α_i è un fattore di scaling temporale che permette di trasporre l'altezza del grano.

Il segnale totale è:

$s_{total}(t) = \Sigma_i A_i g_i(t)$

dove A_i sono ampiezze individuali dei grani.

La sintesi granulare permette manipolazioni timbriche impossibili con altri metodi, come time-stretching senza cambio di altezza (variando le temporizzazioni t_i mantenendo α_i costante) o cambio di altezza senza time-stretching (variando α_i mantenendo la densità temporale dei grani costante).

La sintesi wavetable utilizza tabelle di forme d'onda pre-calcolate o campionate che sono lette a diverse velocità per produrre altezze diverse. Moderne tecniche di wavetable permettono morfing continuo tra diverse forme d'onda mediante interpolazione, creando evoluzioni timbriche fluide.

Modellazione fisica: simulare la produzione del suono

La modellazione fisica sintetizza il suono simulando direttamente i processi fisici di produzione sonora nello strumento. Invece di manipolare direttamente componenti spettrali o forme d'onda, si simulano le equazioni che governano la vibrazione di corde, colonne d'aria, membrane, ecc.

Per una corda vibrante, l'equazione d'onda digitizzata può essere risolta numericamente. Uno schema semplice (metodo delle differenze finite) discretizza l'equazione:

$\partial^2 y/\partial t^2 = c^2 \, \partial^2 y/\partial x^2$

ottenendo:

$y(x, t+\Delta t) = 2y(x,t) - y(x,t-\Delta t) + (c\Delta t/\Delta x)^2 [y(x+\Delta x,t) - 2y(x,t) + y(x-\Delta x,t)]$

dove Δt e Δx sono gli incrementi temporale e spaziale. Questa equazione è iterata numericamente per simulare l'evoluzione temporale della corda.

Un metodo più efficiente è la tecnica della guida d'onda digitale (digital waveguide), sviluppata da Julius Smith. L'idea fondamentale è rappresentare la soluzione dell'equazione d'onda come somma di un'onda viaggiante verso destra y_r(t) e un'onda viaggiante verso sinistra y_l(t):

$y(x,t) = y_r(t - x/c) + y_l(t + x/c)$

Agli estremi della corda (x = 0 e x = L), le riflessioni sono simulate mediante:

$y_r(t + L/c) = -r_1 \, y_l(t)$ $y_l(t + L/c) = -r_2 \, y_r(t)$

dove r_1 e r_2 sono coefficienti di riflessione (tipicamente ≈ -1 per estremi fissi).

Le onde viaggianti sono rappresentate mediante linee di ritardo digitali (delay lines). Una linea di ritardo di durata τ è implementata con un buffer circolare che memorizza campioni passati e restituisce il campione a tempo t - τ.

Lo smorzamento è modellato mediante filtri passa-basso applicati alle onde riflesse, simulando la perdita di energia preferenziale delle frequenze alte. Un filtro del primo ordine:

H(z) = (1 - g)/(1 - g z^(-1))

con g ≈ 0,9-0,99 fornisce smorzamento appropriato.

L'eccitazione (pizzicamento, percussione, archetto) è simulata iniettando appropriati segnali nelle linee di ritardo agli istanti e posizioni appropriati.

Tecniche di modellazione fisica sono state implementate in sintetizzatori commerciali (Yamaha VL, Korg OASYS) e sono utilizzate estesamente in applicazioni di ricerca e educative.

Analisi mediante trasformata wavelet

Mentre la trasformata di Fourier (STFT) utilizza finestre di lunghezza fissa, la trasformata wavelet utilizza funzioni base (wavelets) che hanno durata adattiva dipendente dalla frequenza analizzata. Questo permette migliore risoluzione temporale per frequenze alte e migliore risoluzione frequenziale per frequenze basse, più coerentemente con le proprietà dell'udito umano.

La trasformata wavelet continua (CWT) è definita come:

W(a, b) = (1/√a) ∫_-∞^(+∞) f(t) ψ*((t-b)/a) dt

dove ψ(t) è la wavelet madre, a è il parametro di scala (inversamente proporzionale alla frequenza), b è il parametro di posizione temporale, e * denota il coniugato complesso.

Wavelet comuni includono:

Wavelet di Morlet: ψ(t) = e^(-t²/2) e^(iω₀t)

che è una sinusoide modulata da una Gaussiana.

Wavelet di Haar: ψ(t) = 1 per 0 ≤ t < 1/2, ψ(t) = -1 per 1/2 ≤ t < 1, ψ(t) = 0 altrimenti

la wavelet più semplice.

Wavelet di Daubechies: una famiglia di wavelets ortogonali con diverse proprietà di regolarità.

La trasformata wavelet è particolarmente utile per analizzare transitori rapidi in segnali musicali, poiché le wavelet a scala piccola (alta frequenza) hanno durata breve e possono localizzare temporalmente eventi rapidi, mentre wavelet a scala grande (bassa frequenza) hanno durata lunga fornendo risoluzione frequenziale fine per componenti lentamente varianti.

La CWT produce una rappresentazione tempo-scala che può essere visualizzata come uno scalogramma, analogo allo spettrogramma ma con risoluzione tempo-frequenza variabile.

Compressione audio e codifica percettiva

La comprensione della percezione spettrale ha permesso lo sviluppo di algoritmi di compressione audio che riducono drasticamente la dimensione dei file audio eliminando informazione percettivamente irrilevante.

Il principio fondamentale è il mascheramento: componenti spettrali che sono mascherati da componenti più forti non contribuiscono alla percezione e possono essere eliminati o quantizzati grossolanamente senza degradazione percettiva.

Il codec MP3 (MPEG-1 Audio Layer 3) implementa questo principio:

1. **Analisi mediante banco di filtri:** Il segnale è decomposto in 32 sottobande frequenziali mediante un banco di filtri polifonici (polyphase filterbank).

2. **Trasformata MDCT:** Ciascuna sottobanda è ulteriormente analizzata mediante la Modified Discrete Cosine Transform:

$$X_k = \Sigma(n=0 \text{ a } N-1)\ x_n \cos[\pi/N\ (n + 0,5 + N/2)(k + 0,5)]$$

producendo coefficienti spettrali con risoluzione variabile.

3. **Modello psicoacustico:** Per ciascuna banda frequenziale, si calcola la soglia di mascheramento basata sulla presenza di componenti forti in bande vicine. Questo utilizza modelli delle bande critiche e curve di mascheramento empiricamente determinate.

4. **Quantizzazione adattiva:** Coefficienti spettrali sotto la soglia di mascheramento sono eliminati. Altri coefficienti sono quantizzati con numero di bit dipendente dal rapporto segnale-maschera.

5. **Codifica Huffman:** I coefficienti quantizzati sono codificati mediante codifica Huffman, che usa codici più brevi per valori più frequenti.

Il risultato è una riduzione tipica del fattore 10-12 (128 kbps invece di 1411 kbps per CD audio) con degradazione percettiva minima se i parametri sono scelti appropriatamente.

Codec più moderni come AAC (Advanced Audio Coding) e Opus utilizzano tecniche più sofisticate includendo:

- Predizione temporale (TNS, Temporal Noise Shaping) che sfrutta correlazioni temporali
- Stereo parametrico (Parametric Stereo) che codifica separatamente informazione spaziale
- Switching adattivo tra trasformate diverse (MDCT con diverse lunghezze di finestra)

Questi codec raggiungono trasparenza percettiva (indistinguibilità dall'originale non compresso) a bitrate di 96-128 kbps per musica stereo.

Riconoscimento automatico di strumenti musicali

L'analisi spettrale è fondamentale per il riconoscimento automatico di strumenti musicali, un problema importante in information retrieval musicale.

L'approccio tipico estrae features (caratteristiche) dallo spettro e le utilizza per classificazione mediante machine learning. Features comunemente usate includono:

Centroide spettrale: $SC = (\Sigma_k f_k |X_k|) / (\Sigma_k |X_k|)$

dove X_k sono i coefficienti FFT e f_k le frequenze corrispondenti.

Rolloff spettrale: Frequenza sotto la quale è concentrato il 95% dell'energia spettrale.

Flux spettrale: Variazione frame-per-frame dello spettro: $SF = \Sigma_k (|X_k(t)| - |X_k(t-1)|)^2$

Coefficienti mel-frequency cepstral (MFCC): Rappresentazione compatta dello spettro basata sulla scala mel, che approssima la percezione di altezza. I MFCC sono calcolati come:

1. Calcolare lo spettro di potenza: $P_k = |X_k|^2$
2. Applicare banco di filtri triangolari sulla scala mel
3. Calcolare il logaritmo dell'energia in ciascun filtro
4. Applicare la DCT (Discrete Cosine Transform)

I primi 13-20 coefficienti MFCC catturano efficacemente la forma spettrale (l'inviluppo spettrale) separatamente dalla struttura armonica fine.

Descrittori temporali: Inviluppo temporale, tempi di attacco, centroide temporale, ecc.

Questi features sono utilizzati per addestrare classificatori (Support Vector Machines, Random Forests, Neural Networks). Con features appropriate e dataset di training sufficienti, accuratezze di classificazione del 70-90% sono tipiche per classificazione tra 10-20 strumenti comuni.

Deep learning recente ha migliorato queste performance utilizzando reti neurali convoluzionali (CNN) addestrate direttamente su rappresentazioni tempo-frequenza (spettrogrammi o scalogrammi), evitando la necessità di progettare manualmente features.

Separazione di sorgenti audio

Un problema inverso correlato è la separazione di sorgenti: dato un segnale che contiene multipli strumenti suonati simultaneamente, isolare ciascuno strumento. Questo è estremamente difficile poiché i segnali si sovrappongono nello spettro.

Approcci classici includono:

Filtraggio comb per separare armoniche: Se due strumenti hanno fondamentali diverse f_1 e f_2, le loro armoniche sono a n_1f_1 e n_2f_2. Se f_1/f_2 è irrazionale, le armoniche non coincidono (eccetto accidentalmente) e possono essere separate assegnando ciascun picco spettrale alla fondamentale più vicina.

Non-negative Matrix Factorization (NMF): Lo spettrogramma di magnitudine V (matrice tempo-frequenza non negativa) è fattorizzato come:

$V \approx W H$

dove W è una matrice di templates spettrali e H è una matrice di attivazioni temporali. Ciascuna colonna di W rappresenta lo spettro caratteristico di una sorgente, e ciascuna riga di H rappresenta quando quella sorgente è attiva.

Deep learning: Reti neurali profonde, particolarmente reti ricorrenti (LSTM) e reti U-Net, sono addestrate end-to-end per mappare spettrogrammi di miscele a spettrogrammi di sorgenti individuali.

La separazione sorgenti rimane difficile, ma progressi recenti hanno permesso applicazioni pratiche come rimozione vocale per generare tracce karaoke, estrazione di linea melodica principale, e rimasterizzazione di vecchie registrazioni.

Applicazioni in etnomusicologia e musicologia computazionale

L'analisi spettrale ha applicazioni importanti nella ricerca musicologica:

Analisi di tradizioni musicali non-occidentali: Molte tradizioni utilizzano scale non equaltempered. L'analisi spettrale precisa permette di misurare gli intervalli effettivamente utilizzati. Per esempio, studi sulla musica gamelan indonesiana hanno rivelato scale con deviazioni sistematiche dall'equal temperament.

Studio dell'intonazione storica: Analisi di registrazioni antiche permette di ricostruire prassi di intonazione e temperamento. Per esempio, analisi di registrazioni di pianoforti del diciannovesimo secolo hanno rivelato uso di temperamenti "stiracchiati" diversi dall'equal temperament moderno.

Autenticazione di strumenti storici: Lo spettro caratteristico di violini Stradivari è stato studiato estensamente nel tentativo di comprendere quali proprietà fisiche contribuiscono alla loro qualità sonora leggendaria.

Sintesi vocale e riconoscimento del parlato

La decomposizione sorgente-filtro della produzione vocale (eccitazione glottale + risonanze del tratto vocale) è fondamentale per sintesi e riconoscimento del parlato.

Sintesi vocale concatenativa: Frammenti (diphones, triphones) di parlato registrato sono concatenati. Modifiche di prosodia (intonazione, durata) sono ottenute mediante algoritmi di time-stretching e pitch-shifting basati su analisi spettrale.

Sintesi vocale parametrica: Modelli del tratto vocale generano parlato specificando parametri di formanti, frequenza fondamentale, e caratteristiche di eccitazione. Recentemente, sintesi neurale basata su WaveNet e altri modelli generativi hanno prodotto qualità quasi indistinguibile dal parlato umano.

Riconoscimento del parlato: MFCC estratti da finestre successive del segnale sono utilizzati come features per Hidden Markov Models o Deep Neural Networks che riconoscono fonemi e parole.

Prospettive future: intelligenza artificiale e elaborazione neuromorfiche

Sviluppi recenti in deep learning hanno aperto nuove possibilità:

Reti neurali generative (GANs, VAEs) per sintesi audio: Modelli come WaveGAN generano forme d'onda audio apprendendo distribuzioni di probabilità da grandi dataset.

Neural audio synthesis: Modelli come NSynth (Neural Synthesizer) di Google interpolano tra timbri di strumenti diversi nello spazio latente appreso, creando suoni ibridi impossibili con sintesi tradizionale.

Elaborazione audio neuromorfiche: Hardware specializzato che emula proprietà di reti neurali biologiche potrebbe permettere elaborazione audio a bassissima potenza per dispositivi portatili.

Conclusione: dall'astrazione matematica alla percezione musicale

Il percorso che abbiamo tracciato, dalle equazioni fondamentali del moto ondulatorio attraverso il teorema di Fourier fino alla fisiologia della coclea e alle applicazioni moderne di sintesi e analisi, dimostra l'unità profonda tra matematica, fisica, fisiologia e percezione musicale.

La risposta alla domanda iniziale – perché strumenti diversi suonano diversamente anche alla stessa frequenza fondamentale – è ora chiara: il timbro deriva dalla distribuzione spettrale specifica delle armoniche (le loro ampiezze relative), dalle risonanze formantiche del corpo dello strumento, dalla struttura temporale dei transitori, e dall'interazione di tutti questi elementi con il sistema uditivo umano che effettua analisi spettrale attraverso le proprietà meccaniche della coclea.

La trasformata di Fourier non è semplicemente uno strumento matematico astratto ma descrive accuratamente un processo fisico reale: la decomposizione spettrale effettuata dalla membrana basilare nella coclea. Quando percepiamo un suono come avente un particolare timbro, stiamo effettivamente percependo il pattern di attivazione sulla membrana basilare, che è precisamente la rappresentazione fisica della trasformata di Fourier del segnale acustico.

Questa convergenza straordinaria tra matematica pura, processi fisici, meccanismi fisiologici e percezione soggettiva illustra l'unità fondamentale della conoscenza scientifica. Le equazioni differenziali che descrivono le corde vibranti, le funzioni di Bessel che emergono dall'analisi di Fourier, le proprietà meccaniche graduate della membrana basilare: tutti questi

apparentemente distinti fenomeni convergono per creare l'esperienza ricca e complessa che chiamiamo musica.

Le applicazioni pratiche di questa comprensione – dalla compressione audio che permette streaming musicale via internet, alla sintesi di strumenti virtuali indistinguibili da quelli reali, al riconoscimento automatico di musica che permette di identificare canzoni da brevi frammenti – dimostrano che la teoria non è sterile astrazione ma fondamento di tecnologie che trasformano come creiamo, distribuiamo e interagiamo con la musica.

Guardando al futuro, l'integrazione di modelli computazionali sempre più sofisticati con comprensione sempre più profonda della fisiologia uditiva e della cognizione musicale promette ulteriori sviluppi. Modelli di deep learning che imparano rappresentazioni implicite del timbro da vasti dataset potrebbero superare le limitazioni dei modelli basati su features esplicitamente progettate. Interfacce brain-computer potrebbero eventualmente permettere sintesi musicale controllata direttamente dall'attività neurale auditiva, bypassando interamente strumenti fisici.

Tuttavia, indipendentemente da quanto avanzate diventino le tecnologie, la teoria di Fourier rimarrà fondamentale. Ogni segnale temporale può essere rappresentato nel dominio frequenziale, ogni processo lineare può essere caratterizzato dalla sua risposta in frequenza, e l'orecchio umano opererà sempre come analizzatore spettrale meccanico. Questi fatti matematici e fisici sono invarianti indipendentemente dalle tecnologie utilizzate per produrre, elaborare o analizzare suoni.

La bellezza della teoria di Fourier risiede nella sua universalità e inevitabilità. Non è una costruzione arbitraria ma una conseguenza necessaria della natura lineare dell'equazione d'onda e della completezza delle funzioni trigonometriche come base per lo spazio delle funzioni. In questo senso, la decomposizione spettrale non è semplicemente un modo conveniente di analizzare suoni ma riflette una struttura intrinseca della realtà fisica.

Questa profonda connessione tra struttura matematica e fenomeno fisico, mediata attraverso la percezione biologica per produrre esperienza estetica, costituisce uno degli esempi più eleganti di come la matematica illumini il mondo naturale. Dalle vibrazioni di una corda alla percezione di una melodia, il filo conduttore è la decomposizione in componenti armoniche – un'idea che, concepita originariamente come astrazione matematica, si è rivelata essere la chiave per comprendere uno degli aspetti più immediati e universali dell'esperienza umana: il suono della musica.

Bibliografia tecnica essenziale

Matematica e analisi di Fourier:

Bracewell, R. N. *The Fourier Transform and Its Applications*. McGraw-Hill, 2000.

Oppenheim, A. V., Willsky, A. S., Nawab, S. H. *Signals and Systems*. Prentice Hall, 1996.

Acustica fisica e strumentale:

Fletcher, N. H., Rossing, T. D. *The Physics of Musical Instruments*. Springer, 2008.

Benade, A. H. *Fundamentals of Musical Acoustics*. Dover, 1990.

Chaigne, A., Kergomard, J. *Acoustics of Musical Instruments*. Springer, 2016.

Psicoacustica e percezione:

Moore, B. C. J. *An Introduction to the Psychology of Hearing*. Brill, 2012.

Rossing, T. D., Moore, R. F., Wheeler, P. A. *The Science of Sound*. Pearson, 2013.

Fastl, H., Zwicker, E. *Psychoacoustics: Facts and Models*. Springer, 2007.

Elaborazione digitale del segnale audio:

Zölzer, U. *Digital Audio Signal Processing*. Wiley, 2008.

Smith, J. O. *Spectral Audio Signal Processing*. W3K Publishing, 2011. (disponibile online)

Puckette, M. *The Theory and Technique of Electronic Music*. World Scientific, 2007.

Sintesi e analisi computazionale:

Roads, C. *The Computer Music Tutorial*. MIT Press, 1996.

Serra, X. *Musical Sound Modeling with Sinusoids plus Noise*. In "Musical Signal Processing", 1997.

Applicazioni in musicologia computazionale:

Müller, M. *Fundamentals of Music Processing*. Springer, 2015.

Lerch, A. *An Introduction to Audio Content Analysis*. Wiley-IEEE Press, 2012.